Review on Customer Segmentation Technique on Ecommerce

Article *in* Journal of Computational and Theoretical Nanoscience · October 2016

DOI: 10.1166/asl.2016.7985

CITATIONS
4

READS
4,818

4 authors:

Juni Nurma Sari
Politeknik Caltex Riau
8 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Lukito Nugroho
Universitas Gadjah Mada
179 PUBLICATIONS **733** CITATIONS
SEE PROFILE

Ridi Ferdiana
Universitas Gadjah Mada
111 PUBLICATIONS **200** CITATIONS
SEE PROFILE

Paulus Insap Santosa
Universitas Gadjah Mada
183 PUBLICATIONS **816** CITATIONS
SEE PROFILE

Advanced Science Letters
Vol. 4, 400–407, 2011

Review on Customer Segmentation Technique on Ecommerce

Juni Nurma Sari[1,2], Lukito Edi Nugroho[1],Ridi Ferdiana[1] ,P. Insap Santosa[1]

[1]Department of Electrical Engineering and Information Technology, University of Gadjah Mada, Jogjakarta, Indonesia
[2]Department of Informatics Technology, Polytechnic Caltex Riau, Pekanbaru, Indonesia

Ecommerce transactions are no longer a new thing. Many people shop with ecommerce and many companies use ecommerce to promote and to sell their products. Because of that, overloading information appears on the customers' side. Overloading information occurs when customers get too much information about a product then feel confused. Personalization will become a solution to overloading problem. In marketing, personalization technique can be used to get potential customers in a case to boost sales. The potential customer is obtained from customer segmentation or market segmentation. This paper will review customer segmentation using data, methods and process from a customer segmentation research.The data for customer segmentation were divided into internal data and external data. Customer profile and purchase history were treated as the internal data while server log, cookies, and survey data were as the external data. These data can be processed using one of several methods: Business Rule, Magento, Customer Profiling, Quantile Membership, RFM Cell Classification Grouping, Supervised Clustering, Customer Likeness Clustering, Purchase Affinity Clustering and Unsupervised Clustering. In this paper, those methods were classified into Simple technique, RFM technique, Target technique, and Unsupervised technique and the process was generalized in determining the business objective, collecting data, data preparation, variable analysis, data processing, and performance evaluation. Customer behavior in accessing ecommerce when viewing a product on ecommerce was recorded in server log with time. Duration when seeing the product can be used as customer interest in the product so that it can be used as a variable in customer segmentation.

Keywords: Ecommerce, Customer Segmentation, Personalization

1. INTRODUCTION

Ecommerce development began when the internet is growing and growing until today, especially in B2C ecommerce (Business to Customer). When shopping use ecommerce, a user finds it easy and faster. The ease of using ecommerce encourages customers to buy using ecommerce. With these conditions the problem that comes up is the overloading information because of many products offered by ecommerce[1]. Overloaded information can be overcome by an implementation of personalization in ecommerce services such as providing product recommendation, links recommendation, ads or text and graphics that correspond to the users' characteristics and needs[2]. In addition to solving the problem of overloaded information, personalized services in ecommerce can maintain customer loyalty of existing customer[3], getting new customers by providing service to customers in accordance with their needs and characteristics. It will generate more profits for the company. Before the personalization is implemented, customer segmentation should be conducted because the result from customer segmentation process will be used as inputs to personalize ecommerce services, resulting in dynamic personalization ecommerce services based on current customer conditions.

Customer segmentation is currently performed by processing customer database, i.e. demographic data or purchase history. Several researchers discuss the customer segmentation method on their papers, such as Magento[4], who used several variables to perform customer segmentation, namely transaction variable, product variable, geographic variable, hobbies variable and page viewed variable; Baer[5] and Colica[6] discuss customer segmentation methods of Business Rule, Quantile membership, Supervised Clustering, Unsupervised Clustering, Customer Profiling, RFM Cell Classification Grouping, Customer Likeness Clustering and Purchase Affinity Clustering. Some of these methods have similarity. Other researchers discuss the implementation of customer segmentation. This paper will classify customer segmentation methods based on data processing.

*Email Address: juni.s3te14@mail.ugm.ac.id

1Adv. Sci. Lett. Vol. 4, No. 2, 2011 1936-6612/2011/4/400/008 doi:10.1166/asl.2011.1261

2.CUSTOMER SEGMENTATION

In marketing, one way to increase profits is to communicate with customers to determine customer wishes[5]. Communication is built according to the characteristics of the customer. Communication is very difficult to create using personal approaches. So it is necessary to divide customers into groups that have the same characteristics, and this is called customer segmentation. Schneider[7] also called market segmentation that divides potential customers into a group. Magento[4], an ecommerce platform, in its ebook mentions that customer segmentation is an activity to divide customers into groups that have the same characteristics. Customer segmentation has several benefits: it enables us to match between the customer and an offer of similar products; it changes the way we communicate with the customer based on customer data; it identifies the most profitable customers; and it enables us to update the products and services to meet customer needs. Baer[5] states that customer segmentation is the activity to categorize or to classify an item or subject to a group that has been identified to have in common. In his research, Baer discusses Customer Segmentation Intelligence to improve marketing in offering products or services that meet the needs of each customer group. Segmentation according to Collica[6] is the process to categorize or classify an item into a group that has a similarity in characteristic and in CRM (Customer Relationship Management) segmentation is used to classify customer based on some similarities by segmenting the records of customer database. This chapter will discuss the customer data for customer segmentation, customer segmentation methods and customer segmentation process and then the methods will be classified based on data processing.

A. Data for Customer Segmentation

Customer segmentation requires customer data from various sources. Magento[4] categorizes the data into internal data and external Data. Customer registration, customer profile, and purchase history are the internal data obtained from the database of an ecommerce. While external data are census data, media browsing, surveys and market search, cookies, web and social media analysis. Information about customer lifestyle, attitude, activity and shopping preferences are obtainable through surveys and market search and social media. Browsing history can be seen from server log or cookies. Baer[5] in his research, Customer Segmentation Intelligence, uses internal data by looking the demographic data from customer profile and purchase history. Likewise, Colica[6], uses the customer database and purchase history on customer segmentation methods.

B. Methods of Customer Segmentation

Customer segmentation can be performed using various approaches. Theoretically, Schneider[7] divides customer segmentation methods into geographic, demographic, psychographic, behavioral/occasion, usage-based market segmentation. Geographic segmentation is based on location. Demographic segmentation is based on age, gender, family size, income, education, religion or ethnic. Psychographic segmentation is based on social class, personality or their approach to living. Behavioral segmentation is based on customer behavior but when customer behavior occurs in specific time or occasion, Schneider called it Occasion segmentation. Usage-based Market segmentation is based on behavior pattern of each visitor, which includes a set of categories of customer namely browser, buyer and shopper. Browsers are visitors that just browse a site; buyers are visitors that make a purchase; and shoppers are customers that want to buy, but want to read product reviews and the list of features before buying.

Almost the same with Schneider, Magento divides customer segmentation methods into Profit Potential, Past Purchase, Demographic, Psychographic and Behavior. In Magento[4] there are several variables used:

1) *Profit Potential:* using variable transaction frequency, date of last purchase, average order value, customer lifetime value.
2) *Past Purchases:* using the variable of product type/attribute, product price, payment/shipping method used, product benefit sought (price, quality, prestige), product satisfaction.
3) *Demographic:* using the variable of geographic location (city state, country, region), age, gender, household size, income, occupation, education, ethnicity, browsing device (laptop, PC, tablet, smartphone) and type (vendor and model), traffic source (organic search, banner link, referral site).
4) *Psychographic:* using the variable of hobbies and interest, leisure and recreational activity, affiliations (religious, professional, cultural, political, institutional), personal traits (social vs. private; modern vs. traditional; spontaneous vs. cautious).
5) *Behavior:* using the variable of pages viewed, responses to offers and promotions, participation in reward programs, channel management.

Magento also performed an analysis of purchase history to get the best customer, unprofitable customers, potential customer profit. Best customer is when the customer is a frequent shopper and a repeated customer, with high average order value, low return, providing review and response customer. Unprofitable customer when the customer has high rate product return, low average order value, high rate customer service calls, wants the lowest price. Potential customer profit is determined by counting customer lifetime values.

Baer[5] segments customer using business rules method, quantile membership method, supervised clustering with decision tree method and unsupervised clustering method using k-means algorithm. Demography

data and purchase pattern are used to segment costumers. Here are Baer customer segmentation methods:

1) *Bussiness Rule:* in this method, customers are grouped into specific groups based on a predetermined class, such as:
   a) Grouping based on demographic data, such as age, gender, income and education, etc. This method has similarity with Magento and Schneider.
   b) Grouping based on customer interaction with the company based on data purchase pattern such as the type of product or service provided or RFM data, where R is Recency (when customer last shopped), F is Frequency (how often the customer shops) and M is Monetary (how much the customer spends)

   According to Baer, the lack of business rule does not reflect the actual customer behavior and a segment similar to another segment.
2) *Quantile Membership,* this method uses data Recency, Frequency, and Monetary. Here is the quantile membership methods:
   a) Recency divided into five groups of intervals, for example, starting from 0 days up to 730 days then classify it with label A until E, where A is very valuable customer and E is low-value customer. Also with Frequency and Monetary. When 3 RFM is combined, there is label AAA until EEE.
   b) Map two components of RFM to a table.
   c) Divided into two groups A, B with the classification most valuable customer and two groups D, E to the classification of least valuable customer. C is average value customer.
   d) The result can be inferred for example good frequency (A or B), good monetary (A or B) but poor recency (D or E), and then the advice that given is upgrade the promotion strategy to make the old customer come back
3) *Supervised Clustering with decision tree:* this method uses a specific target, or dependent variable and target would predict differences in independent variables (input). Data utilized in this method is previous purchase pattern and customer demographic.The algorithm that used is decision tree with the target on their nodes. According to Baer, although this method connects the target with the other customer attributes, it shows only one aspect of customer behavior.
4) *Unsupervised Clustering*: this method uses any number of customer attributes then measure the similarity among customer, each customer attribute use Euclidean distance[8] (1) then cluster the customer use k-means clustering[9] (2). If the distance is the shortest distance between customer data and cluster, then customer is included in that cluster.

$$\text{Euclidean distance}=\sqrt{(X_A - X_B)^2}+\ldots+(X_A - X_B)^2 \quad (1)$$

$$C(i)= \arg\min \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \quad (2)$$

Colica has several methods are almost the same. Colica has segmentation methods as follows: Customer Profiling, Customer Likeness Clustering, RFM Cell Classification grouping and Purchase Affinity Clustering. In Customer Profiling method, the required information about customer is the fourW's (who, what, where and when) from customer database.It can be done by using a query on the customer database or using the clustering algorithm when the data is huge. Customer likeness clustering method is used in franchise stores to know whether the profits and turnover of each product in each store are similar, then to review other variables such as demographics. Colica also uses a decision tree for simple clustering the same with Baer. Method RFM (Recency Frequency Monetary) Cell Classification Grouping uses three dimensions to classify each customer in one cell after labeling each level of RFM. Colica names it the Segmentation Using Cell-Based Approach. This method is similar to Baer's quantile membership. Another method used by Colica is Purchase Affinity Clustering. This method uses scoring on interesting in certain products then clusters customer database based on that score to get a similar group.

Table 1. Methods of Customer Segmentation

Paper	Method	Data	Advantage	Disadvantage
Magento (2014)	Magento	Demographic, Purchase History, Data Product, Data Media, Data Marketing, Server Log	Have clear variable customer segmentation	There is no data processing for each variable
Baer (2012)	Bussiness Rule	Demographic, Purchase history	Easy to apply, Use database query	Not focus on customer behavior
	Quantile membership	Purchase history	Can process small data, can be used with other data	Good result obtained when determining a good classification
	Supervised Clustering with decision tree	Demographic, Purchase history	Classify customers according to target	Use one variable to cluster
	Unsupervised Clustering	Purchase history	Use any number of customer attributes	Speed of computation depends on k values
Colica (2011)	Customer Profiling	Demographic, Purchase history	use database query if data is small	Not focus on behavior
	Customer Likeness Clustering	Demographic, Purchase History, Data product	classify customers according to the target	Problem arises when there are different unit in record
	RFM Cell Classification Grouping	Purchase history	Efficient three -dimensional mapping	Good result obtained when determining a good classification
	Purchase Affinity Clustering	Purchase history, Data product	know the products most in demand	Spesific to product segmentation

There are some researches that implement customer segmentation methods according to the table above such as Lieberman[10], who uses combination Business Rule, Customer Profiling, Magento to find how much customer spend money monthly on clothing and how many customers visit monthly; Dodwell[11], who uses RFM Analysis to segment email marketing for potential customer; Birant[12], who uses combination RFM Analysis and Data Mining (Classification Rules and Association Rules) to provide better product recommendation; Han[13], who uses Decision tree model to identify high-value customer; Ma[14], who uses Association Rules and Decision Tree to improve customer loyalty, attract new customer and expand the market effectively; Baer[15], who uses Market Base Analysis, K-means Clustering, and Doughnut Clustering to segments customer based on product, and Ezenkwu[9] and Venkatesan[8], who use K-means Clustering to segment customer.

Based on table and researches above, customer segmentation methods can be classified into: Simple technique, because this method uses database query and statistical data; RFM technique, because this method uses RFM analysis; Target technique, because this method must have target to segments customer, for instance, customer segmentation focus on product, focus on purchase; and Unsupervised technique, because this method uses dynamic data. Figure 1 describes Customer Segmentation Classification.

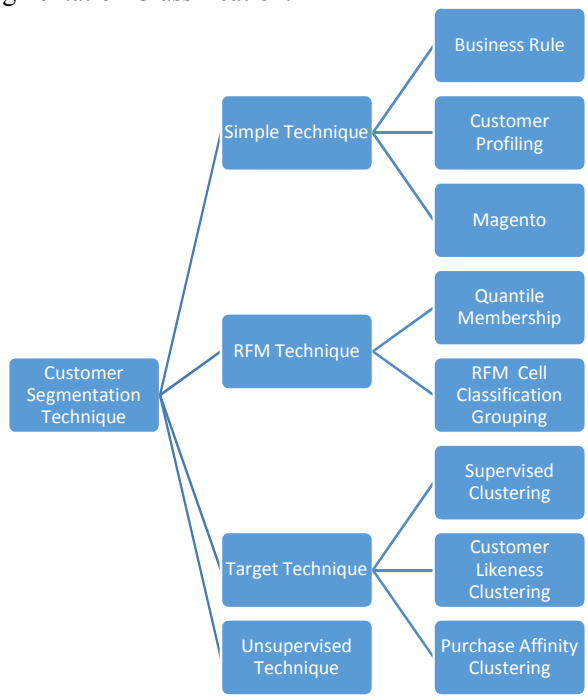

Figure 1. Customer Segmentation Classification

C. *Process of Customer Segmentation*

Customer Segmentation is associated with the business objective. The first step of segmentation is deciding business objective. Chen[16] discusses customer segmentation process begins by determining the business objective such as the identification of high profitable customer groups, improve product for that customer. The next step is collecting the necessary data such as demographic data, transaction data, and promotional data, then determining the method of customer segmentation and standardization measurement. After that, the next step is exploration data by analyzing the statistics and look for relationships between variables. Results of analysis can be used to measure the similarity among the customers using Euclidean distance to measure two points in a multidimensional space where the point is customer data. The cluster is validated by calculating the ratio of the between-cluster variantto within cluster variants(RSQ/1-RSQ).

Process Customer Segmentation on Lieberman[10] research begins with determining the business rule, collecting data spread the questioner, then data processing with logistic regression and waterfall and analyze statistic data. Birant[12] has a more complex process than Lieberman because he combines RFM Analysis and Data Mining to find product recommendation. Birant starts the process of defining the business objective, collecting data, and then data processing with the first method of RFM analysis that uses quantile membership to find customer level of Recency, Frequency, and Monetary. The second method is Clustering with RFM Cell Classification Grouping to find customer segmentation. After segmentation, there is prediction of customer behavior, it uses Association Rule method. Finally, the product recommendation uses Classification method. Process Customer Segmentation on Ma[17] research starts with defining the business objective, choosing variables that relate to purchasing then form data set, finding frequent item set use generalized association rule, cleaning non-interest rule, building tree process, prunning decision tree, extract rules from pruned decision tree in if-then format. Ezenkwu[9] process also starts with determining the business rule, choosing data variable, namely the amount of goods purchased by customer monthly and the average number of customers visiting monthly; the data processing with k-mean clustering which is normalization alongside centroids, initialization step, assignment step and updating step after that performance evaluation. Process of customer segmentation can be simplified into defining business objective, collecting data, data preparation, analyzing variable, data processing, and performance evaluation as describe in figure 2.

3. FUTURE WORK

One of the data used for customer segmentation is customer behavior in accessing ecommerce. Customer behavior data are obtainable from server log. Variables contained in server log are IP address of customer, date, time, HTTP request. Here an example of server log data:

```
05:09:49  GET /detail-item.php?item=ilford-delta-100 HTTP/1.0
05:09:53  GET /detail-item.php?item=ilford-pan-f-50 HTTP/1.0
```

Time shows when a customer accesses page, the difference of time between the customer's visit to the first page and the second page is the duration of customer's visit to the first page. The first data is page detail-item.php with the first product of ilford-delta-100 and the second data is page detail-item.php with the product of ilford-pan-f-50. Knowing data duration, we can determine the user's attention to the product. If the user's attention to the product is in long duration, then the customer has an affinity for product. It can also be used for customer segmentation based on the interest in the product. Such information can be utilized for the promotion of a product. The disadvantage of this method is when customer position isn't in front of computer but server still record the activity, so the solution is using an eye-tracker to record the customer's attention.

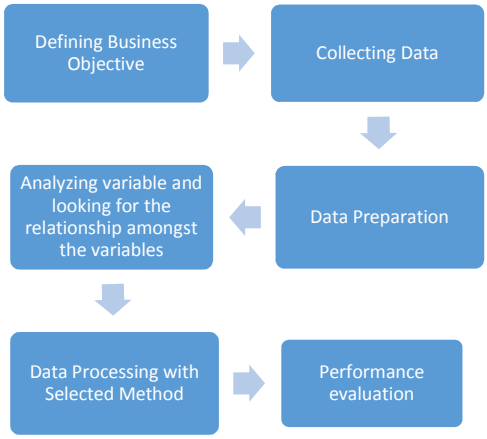

Figure 2. Process of Customer Segmentation

4. CONCLUSION

Customer segmentation is a way to improve communication with the customer, to know the wishes of the customer, customer activity so that appropriate communication can be built. Customer Segmentation needed to get potential customers used to increase profits. Potential customer data can be used to provide service the characteristics of customer including ecommerce services as a media buying and selling online.

This paper discusses several components to do customer segmentation, which is:

- Customer segmentation is an activity to divide customers or item into groups that have the same characteristics.
- Data that needed for customer segmentation are internal data and external data. The internal data include demographic data and data purchase history, while the external data include cookies and server logs. Internal data can be obtained from a database when customer do registration or transactions and external data can be obtained from web server or other source.
- Methods of Customer Segmentation can be classified into Simple technique, RFM technique, Target technique, and Unsupervised technique. On Target technique, researcher focus on one variable, it can be product or purchase. Unsupervised technique was used when clustering process reseacher have many variable
- Process of Customer Segmentation can be simplified into defining business objective, collecting data, data preparation, analyzing variable, data processing, and performance evaluation.

REFERENCES

[1] Al-Qaed F, Sutcliffe A. Adaptive Decision Support System (ADSS) for B2C E-Commerce. *2006 ICEC Eighth Int Conf Electron Commer Proc NEW E-COMMERCE Innov Conqu Curr BARRIERS, Obs LIMITATIONS TO Conduct Success Bus INTERNET*. 2006:492-503.
[2] Mobasher B, Cooley R, Srivastava J. Automatic Personalization Based on Web Usage Mining. *Commun ACM*. 2000;43(8).
[3] Cherna Y, Tzenga G. Measuring Consumer Loyalty of B2C e-Retailing Service by Fuzzy Integral: a FANP-Based Synthetic Model. In: *International Conference on Fuzzy Theory and Its Applications iFUZZY*.; 2012:48-56.
[4] Magento. An Introduction to Customer Segmentation. 2014. info2.magento.com/.../An_Introduction_to_Customer_Segmentation...
[5] Baer D. CSI : Customer Segmentation Intelligence for Increasing Profits. *SAS Glob Forum*. 2012:1-13. http://support.sas.com/resources/papers/proceedings12/103-2012.pdf.
[6] Colica R. Customer Segmentation And Clustering Using SAS Enterprise Minner Part I The Basics. 2011:1-14.
[7] Schneider G. *Electronic Commerce, 9th Edition*.; 2013:643. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
[8] Venkatesan R. Cluster Analysis For Segmentation. 2007.
[9] Ezenkwu CP, Ozuomba S. Application of K-Means Algorithm for Efficient Customer Segmentation : A Strategy for Targeted Customer Services. 2015;4(10):40-44.
[10] Lieberman M. Target " golden egg " consumer to achieve maximum ROI. 2009;(May):50-51.
[11] Dodwell A. Effective Marketing email startegy _ Segmentation RFM. 2015. http://www.sailthru.com/marketing-blog/written-effective-email-marketing-strategies-segmentation-rfm/.
[12] Birant D. Data Mining Using RFM Analysis. *Knowledge-Oriented Appl Data Min*. 2011;(iii):91-108. doi:10.5772/13683.
[13] Hua S, Xiu S, Leung SCH. Expert Systems with Applications Segmentation of telecom customers based on customer value by decision tree model. *Expert Syst Appl*. 2012;39(4):3964-3973. doi:10.1016/j.eswa.2011.09.034.
[14] Ma H. A Study on Customer Segmentation for E-Commerce Using the Generalized Association Rules and Decision Tree. 2015;(December):813-818.
[15] Baer D, Ph D. Product Affinity Segmentation Using The Doughnut Clustering Approach. *Cust Intell SAS Glob Forum 2013*. 2013.
[16] Chen J. Retail Customer Segmentation. 2014;(April).
[17] Chan C, Swatman PMC. Management and business issues for B2B ecommerce implementation. *Proc 35th Annu Hawaii Int Conf Syst Sci*. 2002;00(c):1-11. doi:10.1109/HICSS.2002.994303.