Research on Using Market Segmentation to do Recommendation in E-commerce

Bohan Zhao

School of Information Management and Engineering, Zhejiang University of Finance and Economics, Hangzhou, Zhejiang, China, 310018,
zhaobohanmail@qq.com

ABSTRACT
In the background of the digital period, more and more e-commerce companies have emerged. However, the competition also becomes more intense than before own to a large number of rivals. In this case, it is important to offer the customers recommendation to boost their satisfaction and loyalty. This work will use marketing segmentation as a method to predict the behavior of each customer, and the recommendation system will provide personalized recommendations based on the results. This paper exhibited the process of market segmentation and the K-means algorithm is introduced as the main part to do market segmentation. And the market was segregated based on several characteristics. In addition, the author discussed how to do recommendations by using the result of market segmentation. As a result, two advantages were exhibited: one is boosting the advertising efficiency, another is improving service. Therefore, it could be emphasized that market segmentation completed the recommendation system.

Keywords: *Marketing segmentation, Clustering, K-means algorithm, Recommendation, E-commerce*

1. INTRODUCTION

In the background of the digital period, e-commerce is developing at an astounding speed. In these years, however, the competition among the E-commerce companies are more intense than that used to be because during the last few decades, there are so many companies crowded into this area and it is also a significant reason that traditional commerce is more realizing that digital commerce is a trendy way to get more profit so they transfer their commerce model. Thus, when consumers make a decision, recommendations will play a vital role since they have many other choices. Therefore, it is necessary to offer consumers some accurate recommendations, which means recommending some products they are most likely to buy, only in this way can the consumers buy more things and boost their loyalty. Traditionally, recommendation uses customers' purchasing history information to recommend similar things, which is not reasonable as well as ineffective. In this article, market segmentation will be used to improve the recommendation system. And this paper will introduce the K-means algorithm, which is one of the most useful ways to segment the market. The paper is set out as follows.

First of all, this paper will introduce the concept of market segmentation. Then, the process of segmentation will be discussed. This part mainly uses the method of clustering, like the K-means algorithm. And in the end, this paper will come up with a new recommendation system using the results of market segmentation. This paper hopes to find out what benefits market segmentation can bring to the recommendation and whether it is worth doing. Based on this study, companies in the e-commerce area can get a deeper understanding of the recommendations and come up with more business strategies.

2. MARKET SEGMENTATION

2.1. Concepts of Market Segmentation

Market segmentation is based on the collected data about some characteristics of customers. Therefore, the more data the companies collect, the more accurate the segmentation will be. Companies can use these features to divide their heterogeneous and random customers into several smaller and more homogeneous

subgroups.[1] Through the common features of each group, the system can recommend several similar or related products to the customers in the same group. There is a wide range of characteristics that a company can use to do the segmentation, such as geography, age, gender, income, education and attitudes, values, buyer behavior.[2] For example, people in similar age groups tend to have some analogous interesting things. Like for the students, most of them cannot afford too expensive products, and thus, the company is more likely to recommend some cost-effective things such as some fast-moving consumer goods(FMCG), which are not so durable but at a low cost. In addition, there are four criteria during the process of market segmentation: first of all, the segments must be exist; second of all, the segment must be consistent and repetitive; third of all, the segment must be stable over time; last of all, the segments should be reachable for some people, like through specifically targeted distribution. [3]

2.2. Importance for E-commerce Companies

The importance of market segmentation for E-commerce companies can mainly be introduced from two aspects. [4] First of all, by segmenting a company's target market into several segmented groups, instead of analyzing each customer individually, marketers can use their time, money, or other resources more effectively. As they can use the similarities of a group of people and solve similar situations rather than targeting the consumers on an individual level.

Also, market segmentation reduces the risk of unsuccessful and inefficient decisions in commerce. If a company analyzes one customer individually, it is inevitable to do some ineffective campaign because of the fortuity of the sample. When people are divided by some key characteristics, it is more likely to be successful if they were to create a generic campaign.

3. PROCESS OF SEGMENTATION

3.1. Introduction to Sample

This paper uses open data and code from Kaggle.com.[5] The data includes 5 columns and 200 different customers. This data-set contains 4 key characteristics,'gender', 'age', 'Annual income', and 'Spending score'.

3.2. K-means algorithm

3.2.1. Principles

K-means is one of the oldest and most widely used clustering algorithms. K-means algorithm starts with selecting k objects as the centers and k is given by the users before clustering. This sample uses Elbow Curve to decide the value of k. Using an iteration process, the remaining objects will be assigned to their nearest centers.[6] After the objects are all assigned to a center, the cluster centers will be reconsidered. The new cluster centers have calculated a set of k means P = { $c_1, c_2, c_3 ..., c_k$}. For example, the mean of a set of single-valued objects in cluster i with m points is defined as m_i. The formula of m_i is shown as Eq.(1).

$$m_i = \frac{1}{m}\sum_{j=1}^{m} p_{i,j} \qquad (1)$$

3.2.2. Advantages & Disadvantages

Although K-means is one of the oldest clustering algorithms, there are still some advantages compared with other algorithms. First of all, it is simple to use and the model is easy to debug. Also, it is robust and highly efficient. Last of all, a variety of data types can suit this algorithm so that it can be widely used. [7]

However, it also has some limitations. That is, it is difficult to handle with some complicated data-set, such as a large-scale, high-dimensional one. Also, it is rather sensitive to outliers since cluster centers will be affected significantly by these outliers, which will affect the following iteration process.

3.3. Characteristics & Visualization

3.3.1. Gender

Across some specific products, men and women have different shopping preferences. In this case, gender should be one of the elements that we should take into consideration. In this sample, the ratio of men to women is shown by a pie chart as Fig(1).This chart shows that among the collected 200 samples, 44% are male and 56% are female.

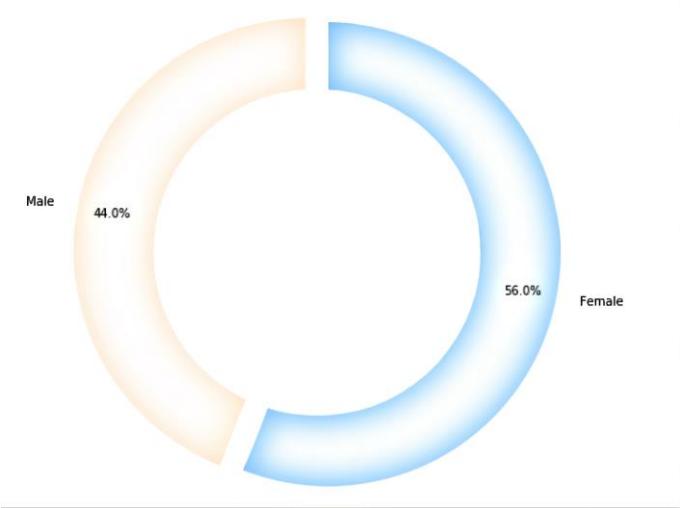

Figure 1. Pie chart of the ratio of men to women [5]

3.3.2. Age

People in similar age groups tend to have similar interests so it plays a significant role in the recommendation. The sample counts the number of each age and uses this as a source to make a bar chart. The chart is shown in Fig(2). This bar chart reveals that in this e-commerce area, the age of the customers ranges from 18 to 70. A company can separate these into several groups. For instance, they can be divided into Young Adults (age 18-30), Early Middle Aged(age 30-40), Late Middle aged(age 40-60), Senior(age 60-70). In this case, it has been revealed that in this e-commerce company's target market, it is more significant to come up with some strategies aiming at Young Adults and Early Middle Aged(age 18-40).

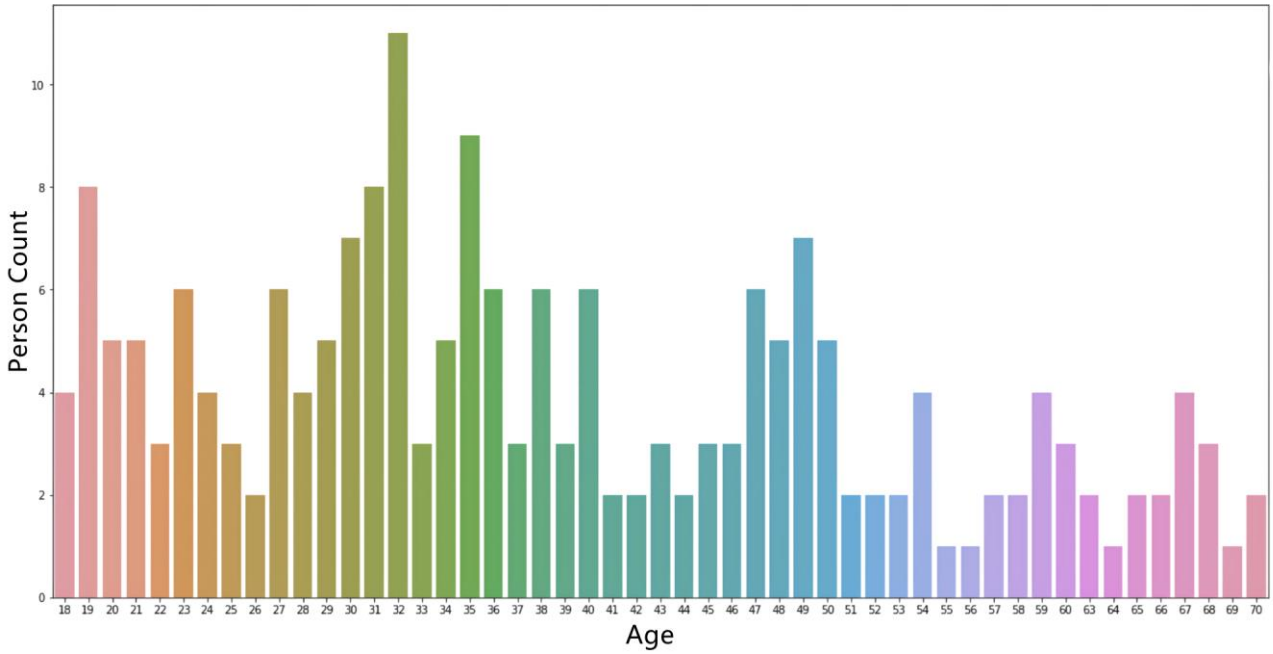

Figure 2. Counting number of each specific age from 200 samples in e-commerce area [5]

3.3.3.Annual income & Spending score

For one company in the e-commerce area, customers' annual income and their spending score are two of the most important characteristics to portrait a customer. These two characteristics can reflect one customer's purchase power and their buying habits, like whether they are more willing to buy luxuries or something more cost-effective. In this sample,

K-means-a specific method to do clustering-is used to do the segmentation.

As mentioned above, the value of k should be determined before the beginning of clustering. In this sample, the author uses Elbow Curve as a method. Elbow Curve is shown in Fig(3).

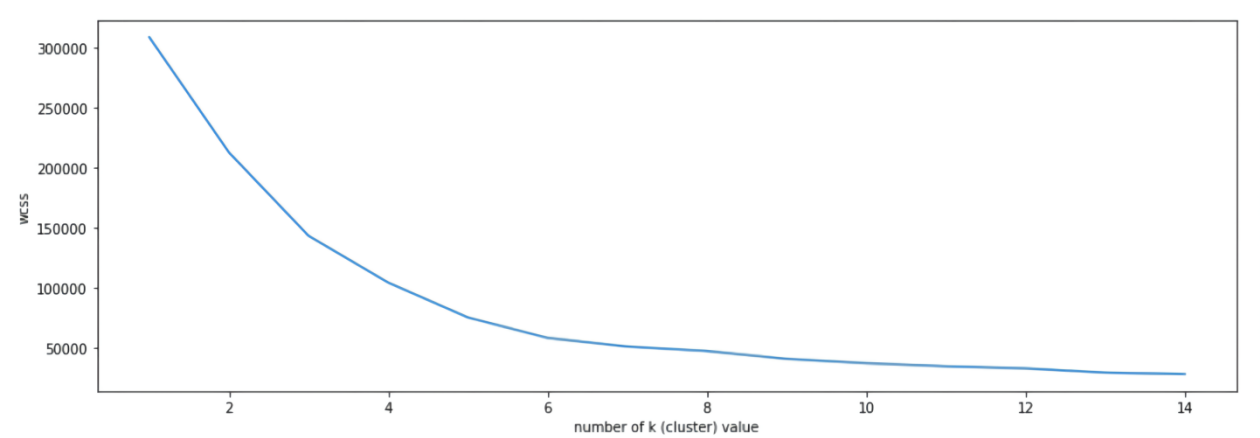

Figure 3. Elbow Curve [5]

It can be seen that after the k value equals 5, the slope of the curve becomes relatively moderate. Therefore, the value of k should take the number equals 5.

After deciding the k value, a k-means model should be created. In the end, 200 samples can be broadly separated into 5 groups according to their annual income and spending score. The scatter plot is shown in Fig(4).

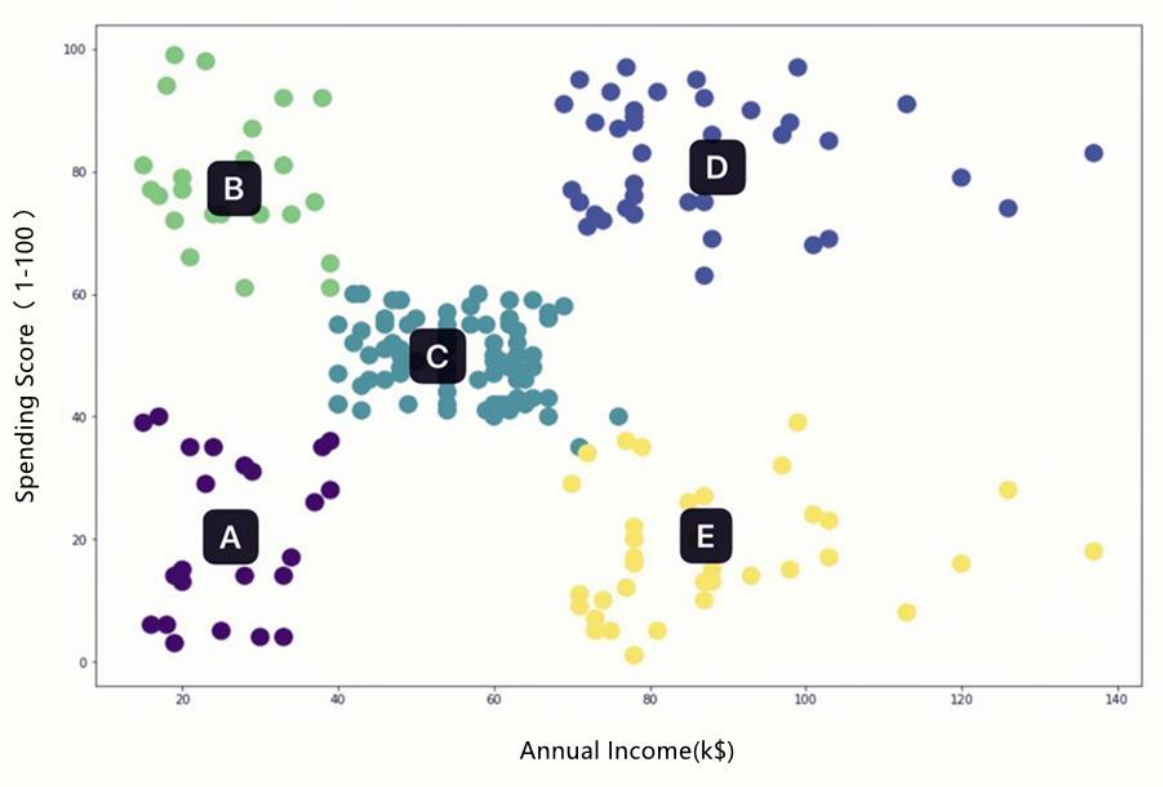

Figure 4. Result of iterating by K-means algorithm (k = 5) [5]

As can be seen in figure 4, clustering of the customers can be named sequentially as A, B, C, D, E. Group A includes people with low income and low purchasing power. Group B includes people with low income but has high purchasing willingness. People in group C have both average income and spending scores. People in group D have high income and high purchasing willingness, while people in group E have high income but they don't buy many goods. An e-commerce company should use these characteristics to portray a customer's figure. In addition, the company should make use of the result of segmentation to do personalized customization. Discussion on how to use the result of segmentation will be discussed as follows.

4. PERSONALIZED RECOMMENDATION

The traditional recommendation in e-commerce is a method using customers' previous purchasing information to learn their preference individually and recommend the product the customers might prefer. For example, if one bought a pair of sunglasses from an e-commerce website, this person will receive numerous advertisements about sunglasses. Although these purchase history can reflect customers' current purchase intention, it is not so reasonable because customers probably will not buy another one, at least in a short period. In this case, a personalized recommendation system using market segmentation is necessary. The utilities of such a recommendation system will be discussed as follows.

4.1. Advertising Effective

The personalized recommendation is potent because after clustering, people in one cluster have similar interests or purchasing power, and they may need similar kinds of products. In this case, if one buys a product on an e-commerce website, this company will advertise similar products to the people in the same cluster. In this way, the advertisement will be more successful and more targeted.[8]

It can be seen in the result of the K-means clustering algorithm that customers in group A have lower income and lower spending scores. There is no doubt that most of the people in cluster A are more likely to buy cost-effective goods, which may also attract others in cluster A to buy the same or similar products. What's more, customers belonging to cluster D are more likely to buy luxuries or high quality products, and thus it is a good idea to advertise such high-quality products to consumers in the same cluster because these customers may have similar elegant tastes. As for the people in cluster C, the annual income is $40k - $60k and the spending score is 40-60, which means they have average income and an average spending score. In this case, they may compare the products and select the most cost-effective things. And what should the e-commerce companies do is recommend such things they want and at the same time, use the fancy advertisement to appeal to them to buy something more expensive. Due to the high income with low expenditure scores, the people in cluster E are the company's main targets because they have great spending potential. Most of the time, they just don't know what to buy so the main task for the company is attracting them by giving the advertisement to introduce the reason why they should buy their product. And for cluster B, with lower income but a high spending score, their consuming behavior has less continuous but these customers are willing to spend money on the things that out of their capabilities, enabling the companies to extract a large amount of profit. Thus, the company still cannot lose them.[9]

As mentioned above, using this personalized recommendation will improve the efficiency and success of advertising.

4.2. Improving Service

In the e-commerce area, one of the most serious limitations is service. The definition of e-commerce service is an e-commerce company that provides facilities or assistance to their customers from the process of making online purchase decisions to resolving issues after their purchase. [10] However, traditionally, the customers were always treated individually. Therefore, the customer service was occupied most of the time. And also, the customer service had little time to listen to customers' requirements carefully. As mentioned above, the customers would be unsatisfied and these problems would undermine the company's profit.

However, after clustering, customers in the same cluster may face similar problems owe to their similar situations and purchase habits. In this case, the e-commerce company can separate the customer service into several groups corresponding to each cluster of the customer and only serve a specific cluster. This method can exert a positive impact on the efficiency of service.

Cluster A will be assigned a small number of service personnel because the companies have the least interest in them. Cluster D and E should be assigned the best service group because they will bring the most profit. Although cluster E has a low spending score, they have the potential to be regular customer if the company provide them with comfortable services, and add new facilities to meet their requirements. And as for cluster B and C, the company can just provide average services.

5. CONCLUSION

This paper studies some advantages of recommendation systems in e-commerce by using market segmentation compared to the traditional recommendation methods. During the research on market segmentation, this paper focuses on the K-means algorithm. To show the results, two charts-visualization of gender and age are exhibited to show the easiest way to do the segmentation. At the same time, the result of the K-means algorithm shows that all customers are grouped into 5 clusters.

As for doing recommendations after market segmentation, there are 2 main advantages compared to the traditional way. First of all, it will boost the efficiency of advertising. In addition, it can improve companies' service offering to the customers in e-commerce.

Nevertheless, there are still some limitations. There are varieties of algorithms to do market segmentation besides K-means. Also, the K-means algorithm is sensitive to the outliers and it is not enough to take 200 samples to debug a model. These problems need to be researched further in the future.

ACKNOWLEDGMENT

The author would like to thank Prof Coggeshall and TA Yuhui, who have contributed to finishing this paper. In the beginning, they devoted their time and energy to help the author determine a topic and frame the whole passage. And because of that, I can manage to complete this work fluently.

REFERENCES

[1] Dudgeon, P. M. Wedel and W.A. Kamakura, Market Segmentation: Conceptual and Methodological Foundations, 2nd edition, Norwell, MA: Kluwer Academic Publishers, 2000, 382 pp. .J. of Classification 19, 179–182 (2002). https://link.springer.com/article/10.1007/s00357-001-0038-6

[2] Nairn, A., Berthon, P. Creating the Customer: The Influence of Advertising on Consumer Market Segments – Evidence and Ethics. Journal of Business Ethics 42, 83–100 (2003). https://doi.org/10.1023/A:1021620825950

[3] Pa. Green and D.Tull. Research for Marketing Decisions. Prentice Hall, 1978

[4] Schlager T., Christen M. (2021) Market Segmentation. In: Homburg C., Klarmann M., Vomberg A.E. (eds) Handbook of Market Research. Springer, Cham. https://doi.org/10.1007/978-3-319-05542-8_29-1

[5] Market Segmentation with Clustering. 2019. https://www.kaggle.com/ecemboluk/market-segmentation-with-clustering

[6] Li X. (2009) K-Means and K-Medoids. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_545

[7] Wu J. (2012) Cluster Analysis and K-means Clustering: An Introduction. In: Advances in K-means Clustering. Springer Theses (Recognizing Outstanding Ph.D. Research). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29807-3_1

[8] Bandyopadhyay, S., Thakur, S.S. & Mandal, J.K. Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society. Innovations Syst Softw Eng 17, 45–52 (2021). https://doi.org/10.1007/s11334-020-00372-5

[9] Zakharov V.N., Philippov S.A. (2017) Clustering of Goods and User Profiles for Personalizing in E-commerce Recommender Systems Based on Real Implicit Data. In: Kalinichenko L., Kuznetsov S., Manolopoulos Y. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2016. Communications in Computer and Information Science, vol 706. Springer, Cham. https://doi.org/10.1007/978-3-319-57135-5_13

[10] Sobihah, M., Mohamad, M., Mat Ali, N. A., & Wan Ismail, W. Z. (2015). E-Commerce Service Quality on Customer Satisfaction, Belief and Loyalty: A Proposal. Mediterranean Journal of Social Sciences, 6(2), 260. Retrieved from https://www.richtmann.org/journal/index.php/mjss/article/view/5806