Innovations

Exploration of clustering techniques in customer segmentation using machine learning

Tannu Chauhan

Student, Department of Mathematics, Chandigarh University,India

Abstract

The rise of various business competitors has resulted in unprecedented level of competition among competitors for new clients and the retention of existing ones. In recent years, every e -commerce company has placed a greater emphasis on Customer Relationship Management (CRM) in order to deliver better customer service than their competitors. Building good client relationships benefits businesses in terms of profits as well as customer retention and satisfaction. In order to acquire valuable insight, businesses must identify potential customers in the market through mining customer data. The rise in business competition, as well as the availability of vast historical datasets, has driven the extensive use of data mining techniques to find useful and strategic information contained in company databases. Clustering analysis is one of the most efficient ways to identify different client traits. The key to company's growth is to begin with an analysis of customer needs and utilize customer segmentation as a means to identify and analyze various customer groups within the system. This will give varied marketing strategies and satisfaction to different types of customers. Customer segmentation is the most useful ways for studying consumer behavior and categorization in business analytics. Customer segmentation can assist in the precise analysis of customer demographics as well as the improvement of service and business quality. In this study different clustering techniques have been discussed in order to segment customers and apply various marketing techniques accordingly. The importance of customer segmentation as a basic CRM function, as well as the several approaches for segmenting consumers using clustering algorithms is also discussed in this study.

Keywords: 1.Relationship; 2.customer; 3.customer segmentation;4.clustering; 5. algorithm; 6.SAPK+K means; 7.Hierarchical; 8.Affinity propagation.

1. Introduction

These days, everything can be customized. Generally speaking, there is no such thing as a one-size-fits-all solution when it comes to the hiring process. This opens up a lot of room for healthy competition between businesses, which in turn allows them to be more creative when it comes to recruiting and retaining customers. In the face of commodity rivalry, businesses should mine client's assets to attain approached goals for distinct clients as well as supply the facilities they desire [7].

Customer segmentation is the process of categorizing your customers on a number of characteristics (for example grouping customers by age). Customer segmentation is popular since it aids in more successful marketing and sales. Through Customer Segmentation you can gain a better comprehension of your client's desires and needs.

The business impact of doing this is considerably greater and effective because customer segmentation can help you to enhance client lifetime value. This indicates that they will stay longer and spend more money on your product. It's a method for organizations to learn more about their customers. When you understand the differences between consumer segments, it's easier to make strategic decisions about product growth and marketing. It's also a better predictor of

customer behavior, which can help businesses make better judgments. It will not only promote customer loyalty but also increase the customer's value, thereby increasing the customer's lifetime value.

The clustering algorithm is a tool for segmenting customers. The clustering technique is a method used by marketing, sales, and other professionals to understand their clients in terms of both their static demographics (such as age, location, and background) and their dynamic behavior (such as the number and nature of interactions with the brand).Clustering algorithms identify groups that are internally similar and externally dissimilar [8]. Through clustering algorithms, customers with similar needs and behaviors are organized into homogeneous clusters [8]. Clustering method is a category of data mining technique that is commonly employed in the analysis of distribution features in data sets in achieving strategic goals in the organization. [10].

Some of the clustering techniques that are used for customer segmentation include K-Means, Hierarchical Clustering, Density-based Clustering, and Affinity Propagation Clustering. The K-means algorithm can help telecom operators perform customer segmentation and precisely discover consumers' market needs in a number of ways [13, 15]. In addition to K-Means, other clustering algorithms, such as Hierarchical clustering, Density-based clustering, and Affinity propagation clustering, have also been discussed in this paper. Combining clustering algorithms can produce better clustering results than using each target individually [3].

Customers are different, and so are the needs and wants of each individual customer. Clustering techniques, therefore, aim to analyze the behavior, requirements, and wants of different customer types so that the marketing of a company can be directed toward the most profitable and valuable group of customers [3]. There are certain factors that are used for the segmentation of customers. These clustering parameters are classified as geographic, demographic, psychographic and behavioral [9]. Predict future consumer consumption trends using customer data and behavior segmentation, as well as profit market planning for businesses, in order to achieve the target of optimal service allotment of assets and the much valuable design of customer merchandising campaigns. [17].

2. Clustering Techniques on Customer Segmentation

Clustering is the process of dividing the community or set of data points into several groups so that data points in the equivalent group are more likely to each other and unlikely to data points in further set of group [3]. It is essentially a collection of objects based on their likeness and unlikeness. There are different clustering algorithms i.e. K-Means, Hierarchical, Density Based, Affinity Propagation, AP algorithm, SOM(Self Organizing Maps) and many more are available for effective classification.

K-Means and Hierarchical Clustering Techniques were applied on a small dataset and it was witnessed that mixture of these techniques could perform better than the individual models [8]. K-means Clustering was applied on the dataset collected from a mega retail business and K-means outperformed with accuracy of 95% [12]. Hierarchical Agglomerative Clustering technique was quite slow and hardware dependent when implemented in R programming for customer segmentation [5]. A model was developed using K-means clustering and SPSS Tool and K-means gave high accuracy when carried out with different analysis techniques [14]. A RFM Model and PCA-AdaBoost Model were also implemented on the E-commerce dataset available on Kaggle and it was observed that PCA-AdaBoost outperformed [1]. SOM NEURAL NETWORK and C4.5 were used for customer analysis and business intelligence tool, DEA (Data Envelopment Analysis) was used for survey-based profitable customer segmentation system [17]. It has observed that paradigm is shifted to Clustering to Supervised learning [1].

Each clustering algorithm has its own set of rules for determining how similar each observation is to its group. For consumer analysis, K Means is the most extensively used clustering method. Hierarchical clustering requires no initial number of cluster conditions, and is best suited for small to medium-sized datasets. The density-based clustering approach can be utilized to find arbitrary shaped clusters, although it is incompetent when there is a large density difference between data points. The affinity propagation algorithm does not requisite an initial cluster, and the clustering result efficiency of this technique is better, as well as the clustering time, that checks high clustering

efficiency and applicability from small to medium datasets. The following papers have been studied related to clustering techniques and represented in table 1.

Table 1. Summary of Major Findings of Customer Segmentation using Clustering Techniques

S.No.	Ref. No.	Findings	Result
1.	[8]	K-Means and Hierarchical Clustering Techniques were applied on a small dataset and it was witnessed that mixture of these techniques could perform better than the individual models.	K-means
2.	[3]	K-Means, Hierarchical, Density based and Affinity Propagation Clustering Techniques were compared on the basis of some factors.	Each clustering technique has its own benefits and limitations. It was observed that K-Means is better than other techniques for customer analysis.
3.	[7]	K-Means and AP algorithm were combined together and a new algorithm SAPK+K means was developed. Iris and Ionosphere dataset was used for the implementation of this new algorithm.	When combined together, K-Means and the AP algorithm gave better results.
4.	[15]	K-means clustering method was implemented on data obtained from small business customers using the data mining tool KXEN to provide a solution for telecom customer segmentation. The target marketing approach proposed in this research was successful and effective.	K-means technique gave effective results when used with KXEN
5.	[14]	A model was developed using K-means clustering and SPSS Tool was used to create a real-time and online system for a precise supermarket in order to project sales during innumerable yearly cycles. ANOVA analysis was also carried out.	K-means gave high accuracy when different analysis techniques were carried out.
6.	[9]	K-means clustering was applied on Instacart Online Grocery Shopping Dataset 2017. To give relevant information to the customers, for this purpose Singular Value Decomposition technique was used	K-means
7.	[17]	SOM NEURAL NETWORK and C4.5 were used for customer analysis and business intelligence tool, DEA(Data Envelopment Analysis) was used for survey-based profitable customers segmentation system	SOM(Self Organizing Map)
8.	[12]	K-means Clustering was applied on the dataset collected from a mega retail business	K-means with 95% accuracy
9.	[13]	K-Means Algorithm with data mining tool KXEN was applied to telemarketing of telecommunication	Using K-means with data mining tool was successful and effective
10.	[5]	Hierarchical Agglomerative Clustering technique was implemented in R programming on credit card dataset for customer segmentation. As an experiment environment Google cloud platform with hardware framework of 26GB of RAM and 4vCPUs was used.	This algorithm was quite slow and hardware dependent.

11.	[6]	K-Means Clustering was implemented to identify profitable customers in an insurance company.	K-means
12.	[19]	K-means, two-phase algorithm and DBSCAN were implement on a bank dataset	K-means
13.	[2]	K-Means, Hierarchical Clustering techniques with PCA were implemented on credit card transaction history	K-means
14.	[21]	Two-stage Clustering method was used for market segmentation	Genetic Algorithm
15.	[1]	RFM Model and PCA-AdaBoost Model were implemented on the E-commerce dataset available on Kaggle	PCA-AdaBoost

Figure 1. represents the literature review of unsupervised learning, supervised learning and both. From Table 1., it has been observed that 85% unsupervised learning was utilized for customer segmentation while 5% supervised learning and 10% both were utilized.

Fig. 1. Paper Distribution for Clustering Technique
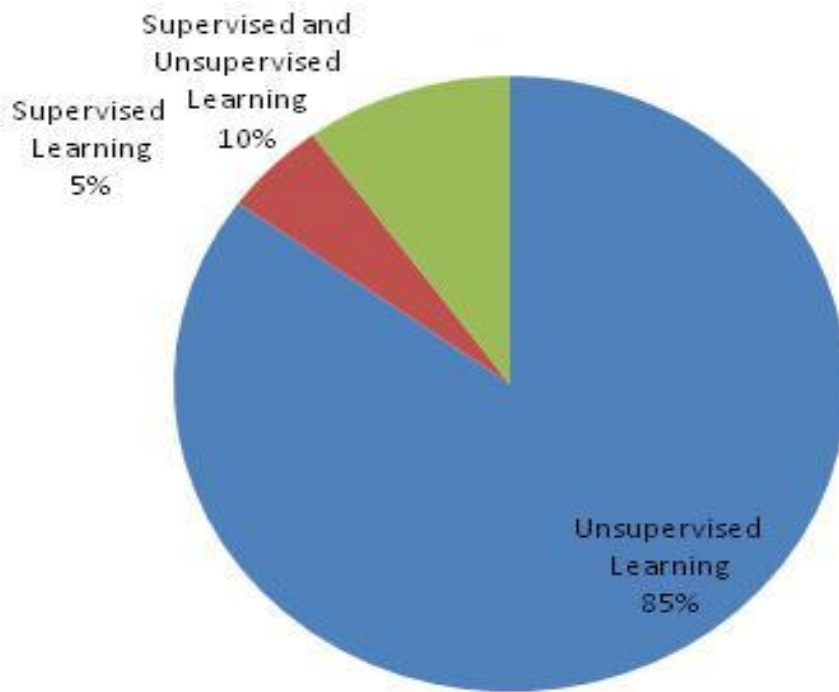
3. Result & Discussion

When working with vast amount of statistics, organizations must employ better clustering algorithms for customer segmentation. These clustering models must be able to effectively process such large amounts of data. Each of the clustering algorithms listed above has its own set of advantages and disadvantages. With the various techniques mentioned above, it was discovered that a hybrid approach of mixing the algorithms can be effective depending on the situation and necessity, and the approach can be applied accordingly. The clustering approach selection process would take a long time to study, carry out, and operate data with an acceptable grasp of the objective and execute the algorithm on requirement basis [3].

It was witnessed that when combined together, K-Means and the AP algorithm gave better results [7]. K-Means with data mining tools was also effective [13]. Some algorithms like Hierarchical Agglomerative was sort of slow and hardware dependent [5]. It was advised to use cloud computing environment in this case.

This work can be extended with Supervised Learning to get the result in transparent and secure manner.

4. Conclusion

In this paper we have surveyed studied current techniques used for customer segmentation. Customer segmentation is important for a company's growth and for a better understanding of your customers. Recognizing the distinct group of clients that boosts earnings would be beneficial to the company. It likewise helps them in supporting client connections and holding clients by carrying out different advertising strategies.

Different clustering techniques can be used with data mining tools for a better prediction. It is clear that model improves accuracy and precision of customer segmentation when different clustering methods were combined together such as K-mean and AP algorithm. Supervised machine learning such as PCA-AdaBoost is also performing well in order to obtain better results in customer segmentation system.

References

1. Wu, Z., Jing, L., Wu, B., & Jin, L. (2022). A PCA-AdaBoost model for E-commerce customer churn prediction. Annals of Operations Research, 1-18.
2. Abdulhafedh, A. (2021). Incorporating k-means, hierarchical clustering and pca in customer segmentation. Journal of City and Development, 3(1), 12-30.
3. Monil, P., Darshan, P., Jecky, R., Vimarsh, C., & Bhatt, B. R. (2020). Customer Segmentation Using Machine Learning. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 8(6), 2104-2108.
4. Du Toit, J., Davimes, R., Mohamed, A., Patel, K., & Nye, J. M. (2016). Customer segmentation using unsupervised learning on daily energy load profiles. Journal of Advances in Information Technology Vol, 7(2), 69-75.
5. Hung, P. D., Lien, N. T. T., & Ngoc, N. D. (2019, March). Customer segmentation using hierarchical agglomerative clustering. In Proceedings of the 2019 2nd International Conference on Information Science and Systems, 33-37.
6. Abolmakarem, S., Abdi, F., & Khalili-Damghani, K. (2016). Insurance customer segmentation using clustering approach. International Journal of Knowledge Engineering and Data Mining, 4(1), 18-39.
7. Deng, Y., & Gao, Q. (2020). A study on e-commerce customer segmentation management based on improved K-means algorithm. Information Systems and e-Business Management, 18(4), 497-510.

8. Tripathi, S., Bhardwaj, A., & Poovammal, E. (2018). Approaches to clustering in customer segmentation. International Journal of Engineering & Technology, 7(3.12), 802-807.

9. Bhade, K., Gulalkari, V., Harwani, N., & Dhage, S. N. (2018, July). A Systematic Approach to Customer Segmentation and Buyer Targeting for Profit Maximization. In 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-6.

10. de Oña, J., de Oña, R., & López, G. (2016). Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation. Transportation, 43(5), 725-747.

11. Hoegele, D., Schmidt, S. L., & Torgler, B. (2016). The importance of key celebrity characteristics for customer segmentation by age and gender: Does beauty matter in professional football?. Review of Managerial Science, 10(3), 601-627.

12. Ezenkwu, C. P., Ozuomba, S., & Kalu, C. (2015). Application of K-Means algorithm for efficient customer segmentation: a strategy for targeted customer services.

13. Ye, L., Qiuru, C., Haixu, X., Yijun, L., & Guangping, Z. (2013). Customer segmentation for telecom with the k -means clustering method. Information Technology Journal, 12(3), 409.

14. Kashwan, K. R., & Velu, C. M. (2013). Customer segmentation using clustering and data mining techniques. International Journal of Computer Theory and Engineering, 5(6), 856.

15. Qiuru, C., Ye, L., Haixu, X., Yijun, L., & Guangping, Z. (2012, August). Telecom customer segmentation based on cluster analysis. In 2012 International Conference on Computer Science and Information Processing (CSIP), 1179-1182.

16. Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. Expert systems with applications, 31(1), 101-107.

17. Lee, J. H., & Park, S. C. (2005). Intelligent profitable customers segmentation system based on business intelligence tools. Expert systems with applications, 29(1), 145-152.

18. Zakrzewska, D., & Murlewski, J. (2005, September). Clustering algorithms for bank customer segmentation. In 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), 197-202.

19. Wu, J., & Lin, Z. (2005, August). Research on customer segmentation model by clustering. In Proceedings of the 7th international conference on Electronic commerce, 316-318.

20. Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. Technology in society, 24(4), 483-502.

21. Kuo, R. J., Ho, L. M., & Hu, C. M. (2002). Cluster analysis in industrial market segmentation through artificial neural network. Computers & Industrial Engineering, 42(2-4), 391-399.

22. Aaker, J. L., Brumbaugh, A. M., & Grier, S. A. (2000). Nontarget markets and viewer distinctiveness: The impact of target marketing on advertising attitudes. Journal of Consumer Psychology, 9(3), 127-140.

tannuchauhan0801@gmail.com