MACHINE LEARNING APPROACHES FOR CUSTOMER SEGMENTATION

Anju.L[1]* , V Sangeetha[2]

Abstract

Nowadays many organizations are giving importance to Customer Relationship Management to increase their profit, customer satisfaction and for retaining the customers. Personalization has a crucial role in business since it provides opportunity for the companies to think creative about how to retain and grab their customers. Customer segmentation is one of the fundamental ways towards personalization. Proper segmentation will help the organizations to make crucial decisions regarding pricing, marketing strategies, product recommendation etc. It is very difficult to conduct this segmentation manually. In this paper the benefits of using Machine Learning in customer segmentation has been described. Clustering technique is one of the methods to segment the customer into different groups. This paper mainly concentrates on various approaches for categorizing the customers.

Keywords - E-commerce, Customer Relationship Management, Customer Segmentation, Machine Learning

I. INTRODUCTION

Segmenting the customers has become an important process in Customer Relationship Management. Success of many businesses depends on the retention and satisfaction of their customers. Most of the e commerce organizations are using different techniques to categorize their customers. The segmentation can be based on many factors like age, purchasing habit, education level etc. The clustering of customers helps the e-commerce organizations to use appropriate marketing strategies.

Department of Computer Science
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

II. CUSTOMER SEGMENTATION

Customer segmentation means that grouping the customers into different groups based on some attributes such that each group will have some kind of similarities among them. Machine learning has made the process of customer segmentation an easy task. Main aim of the customer segmentation in business is that they can identify the group of customers having similar attributes and based on that corresponding marketing strategies can be applied. This helps them to reduce the wastage of money. The following are the advantages of customer segmentation.

* To improve customer satisfaction
* To increase lifetime values of the customer
* To reduce customer churn
* To increase the revenue and sales
* To plan a better use of marketing budget
* To improve customer acquisition

The key business opportunities of customer segmentation are in

* Service or product design
* For promotion of product / service
* Marketing of products
* For making better customer satisfaction
* Budgeting

III. MACHINE LEARNING

Machine learning is a branch of Artificial Intelligence (AI) which allows the computer system to learn from past data. It builds mathematical models by using various algorithms and make predictions with the use of historical data. The application of Machine Learning lies in image recognition, recommender system, email filtering etc.

Machine learning algorithms are divided into several categories.

- Supervised Learning
- Unsupervised Learning
- Semi- Supervised Learning
- Reinforcement Learning

These four classifications are again subdivided into another types.

IV. CLUSTERING IN MACHINE LEARNING

Clustering is also called cluster analysis which is an unsupervised machine learning technique. No prior supervision is provided to the algorithms. Clustering is applied to unlabeled dataset. Clustering algorithms will group the unlabeled dataset into different clusters based on certain attributes. Each object in a cluster will show some kind of similarities and objects in different clusters has less or no similarities among them.

The advantage of computer-based cluster analysis is that manual data analysis is more tedious task and error prone. Since the amount of data is growing rapidly, the processing and updating of data models will be a difficult task. The following are the advantages of using machine learning for customer satisfaction.

- Less processing time.
- More accurate results
- Scalability

Cluster analysis is mainly used for two purposes: utility and for understanding [3]. The main purpose of clustering data is:

- To get useful information from data like detecting anomalies, identification of certain features ,generating hypothesis etc.
- To identify degree of similarity among the organisms, forms, points that comprise the data.
- As a method for organizing data and summarizing data through cluster prototypes [1]

There are mainly two types of Clustering. They are.

- Hard Clustering- Here objects will belong to either a single cluster or none of the cluster.
- Soft Clustering: Here an object can simultaneously belong to more than one cluster.[2]

V. CLUSTERING PARAMETERS

The clustering is performed based on certain parameters called clustering parameters. Clustering parameters can be broadly classified into [1].

-
- Geographic Parameters
- Demographic Parameters
- Psychographic Parameters
- Behavioral Parameters

A. Geographic Parameters

Geographic clustering indicates grouping of customers based on geographical areas. Usually, customers are classified geographically as urban or rural. It can also be based on customer location, city, zip code etc.

B. Psychographic Parameters

Psychographic clustering groups customers based on their lifestyle or personality. That means clustering based on the belief, social status, opinion,values of the customer. This segmentation is different from the behavioral segmentation.

C. Demographic Parameters

Segmenting customers based on their age, education, occupation, religion are the factors considered in demographic clustering.

D. Behavioral Parameters

This focused on the purchasing behavior patterns of the customer, like the attitude of a customer towards a particular brand, product, or a service. Usually, it includes the loyalty or satisfaction of a customer.

VI. CLUSTERING ALGORITHMS

Various clustering algorithms are there. Each clustering algorithm is aimed at set of applications. Different methodologies are specified by each algorithm to specify the closeness of data objects. More than 100 clustering algorithms are proposed and studied.

A. Prototype Based Algorithms

This category of algorithm will learn a prototype for each cluster. Then clusters are formed by data objects around these prototypes. In the case of K -Means [4], and Fuzzy -c-Means (FCM) [5], centroid is taken as prototype of a cluster. A variant of artificial neural network called Self - Organizing Map (SOM) is another prototype-based algorithm, which use a neighborhood function in order to preserve topological properties of a data object. Competitive process is used to train the weight of whole network. Probability distribution function is used to characterize the prototype in Mixture Model (MM), which uses Maximum Likelihood Estimation (MLE) method to estimate unknown parameters [6].

B. Density Based Algorithms

This class of algorithm will take a dense region of data objects as a cluster. This dense region will be surrounded by a low-density region. These categories of algorithms are useful when the noise or outliers are present, or irregular or intertwined clusters are present. Most poplar density based algorithms are DBSCAN [7] and DENCLUE [8].In DBSCAN, data points are divided to create a cluster based on Euclidean Density. In DENCLUE utilizes probably density function based on the kernel functions of each data points and then detect their variance.

C. Graph Based Algorithms

In this method, each data object is represented as a node. Weight of the edges denotes the distance between two nodes. A connected subgraph defines a cluster here.Upto a certain level we can consider Agglomerative hierarchical clustering algorithm(AHC) as graph based algorithm[9].A typical example of graph based algorithm is e Jarvis-Patrick (JP) algorithm. It utilizes shared nearest neighbor for specifying each data object and the clusters are represented by a graph [10].

D. Grid – Based algorithms.

These algorithms will not use the database direct. Usually, statistical methods are used here in order to collect data from database. Performance of these algorithm is proportional to the grid size not the actual data space. One advantage of this algorithm is that it will take less time. Once the grids are formed, the cell density in the grid will be calculated. If the density is below a user specified threshold value, the cell will be discarded. The contiguous groups of dense cells represent the clusters [11]. STING,CLIQUE are examples of grid based algorithms.

D. Hybrid Algorithms

It is possible to combine two or more clustering algorithms together. Such algorithms are called hybrid algorithms. By combining these algorithms, we can overcome the drawback of Single clustering algorithm. One example of hybrid algorithm is Chamelon [10]. It is a combination of graph-based algorithm and AHC. It uses a combination of Graph based algorithm and AHC to get final clusters.

E. Algorithm -Independent Methods

In Consensus clustering, we will combine several clusters into more stable single cluster. This is an iterative process. Consensus clustering is also called cluster ensemble or clustering aggregation. Consensus based clustering algorithm normally include a combination of graph based algorithm, methods based on co- association matrix [12] and also a prototype based clustering method [13,14].CPSA, HGPA and MCLA[11] are some popular graph based algorithms

Criteria	K-Means	Hierarchical Clustering	Density Based Clustering
Speed of Computation	High	Low	Low
Clustering time	Less	More	More
Granularity	Yes	No	Yes
Effect on size of data	Good	Not good	Not Good
Handle Dynamic Data	Yes	No	Yes
Clustering result efficiency	Medium	Low	Medium

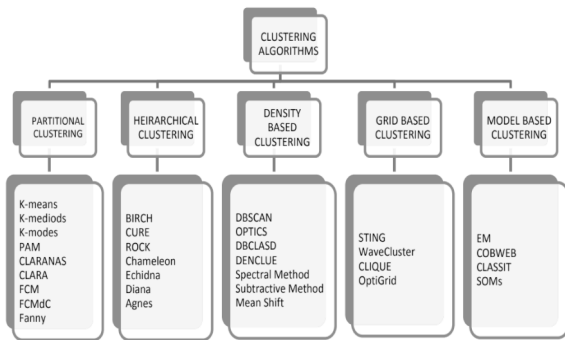Table 1: Comparison of various Clustering Techniques [15]

Figure 1: Categorization of Clustering Algorithms [16]

VII. APPLICATIONS

In various real-world applications, the clustering algorithms are successfully implemented. The following figure will show some among them.

Area of application	Algorithm used
1. In Banking:	
● Anti Money Laundering Regulatory System	– DBSCAN
● Broadcasting warning messages against bank frauds	– K-means, K-means++
● Installing ATMs and e-corners at strategic locations	– DBSCAN
● Long-term bank failure prediction	– Fuzzy Refinement domain adaptation
● Profit maximization and clientele expansion, customer churn management	– Subtractive Clustering Method(SCM), Fuzzy c-means(FCM)
2. In segmentation engines and recommendation engines:	
● Market analysis	– K-means
● Tourist market segmentation	– Fuzzy C-medoids
● Education	– K-means and Hierarchical clustering
● Crime domain documentation	– K-means, k-means++
● Drawing patterns in social media usage, handling all the side-information and metadata within any document, working of web search engines	– k-means, fuzzy C-means and hierarchical clustering algorithms

Fig 2: Application of Cluster Analysis [17,18]

3. In Health Care:	
● Detection of neurovascularization in retinal images	– Multivariate m-Medoids based classifier
● Detection of tumors	– K-means
● Blood oxygen level dependent functional MRI activation detection	– Mean shift algorithm
● Brain image segmentation	– Spectral clustering
● Inhomogeneous medical image segmentation	– Fuzzy clustering
● Medical image analysis	– Spectral clustering
4. In Urban Development:	
● Positioning of towns, building efficient power supplies	– K-means clustering based on Ant Clustering Method
● Setting up amenities such as banks, schools etc.	– DBSCAN
● Transformer fault diagnosis	– k-medoids
● Identifying patterns in online cable-discharge monitoring	– k-means
● Population estimation with satellite imagery	– Expectation Maximization

Fig.3: Application of Cluster Analysis [19,20]

5. In Network analysis:	
● Speed up data transfer by transporting clusters of similar traffic along network at once	– K-means, DBSCAN and AutoClass Algorithms
● Reducing energy consumption in sensors across sensor network	– Hierarchical clustering algorithm
6. In Privacy protection:	
● Responsible and secured data sharing	– k-means and distributed clustering
● Elimination of scam websites	– Combined clustering method
● Behavioral malware clustering	– Single linkage hierarchical clustering
● Anomaly based intrusion detection	– k-means, Opti-Grid Clustering, cluster labeling
7. In Image segmentation:	
● Partitioning image pixels into super-pixels	– FCM, FELICM
● Hierarchical image segmentation	– Correlation clustering method
● Hierarchical pixel clustering	– k-means method
● Medical image analysis	– k-means algorithm

Fig.4: Application of Cluster Analysis [21]

VIII. POPULAR CLUSTERING ALGORITHMS AND ITS APPLICATION AREAS

The following figure will explain some of the popular Clustering Algorithms and the application areas where these algorithms are used.

Most popular algorithms	Areas of application
K-Means	● Broadcasting warning messages against bank frauds
	● Market analysis, crime domain documentation
	● Drawing patterns in social media
	● Web search
	● Detection of tumors
	● Identifying patterns in online cable-discharge monitoring
	● Speed up data transfer
	● Responsible and secured data sharing
	● Medical image analysis

Fig 5: Popular Clustering Algorithm and its Application areas [22]

Hierarchical Clustering Algorithms	• Education • Drawing patterns in social media • Web search engines • Handling side information and metadata within any document • Reducing energy consumption in sensors across sensor network • Behavioral malware clustering • Hierarchical image segmentation
DBSCAN	• Anti Money Laundering Regulatory System • Profit maximization and clientele expansion • Customer churn management • Managing network traffic • Setting up amenities such as banks, schools etc. • Installing ATMs and e-corners at strategic locations
Mean-shift Algorithm	• Blood oxygen level dependent functional MRI activation detection • Image processing • Medical Image analysis
Expectation Maximization	• Population estimation with satellite imagery • Image reconstruction • Auto fill function in web search engines

Fig 6: Popular Clustering Algorithm and its

Application areas [23]

IX. CONCLUSION

For a business, it is not a good method to handle all customers equally. The needs of each customer will be different. Customer segmentation is one of the ways to solve this problem. Advantage of customer segmentation is that it will improve the customer experience and will have a positive influence towards the company revenue. It will help to surpass the competitors and help to get more and more customers. Machine Learning has made the process of customer segmentation an easy job. It provides various algorithm to perform clustering of customers. In this paper some important clustering algorithms, their comparisons and some of the real-world applications where each one will be useful has been described. There is numerous application domain for clustering such as Pattern Recognition, Business Intelligence, Social Science, Artificial Intelligence, Bio Informatics etc. Based on the scenario and based on accuracy and efficiency appropriate algorithms can be selected.

REFERENCES

[1] Shivam Agarwal" Data Mining: Fata Mining Concepts and Techniques" IEEE International Conference on Machine Intelligence Research and Advancement 2013.

[2] M. S. Chen, J. Han, and P. S. Yu, "Data mining: An overview from a database perspective," IEEE Trans. Knowl. Data Eng.,vol. 8, no. 6, pp. 866–883, Dec. 1996.

[3] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," Data Mining Knowl. Discovery, vol. 8, no. 1, pp. 53–87, 2004.

[4] C. C. Aggarwal, Y. Li, J. Wang, and J. Wang, "Frequent pattern mining with uncertain data," in Proc. 15th ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining, 2009, pp. 29–38.

[5] R. Agrawal, T. Imieli_nski, and A. Swami, "Mining association rules between sets of items in large databases," ACM SIGMOD Rec., vol. 22, no. 2, pp. 207–216, 1993.

[6] R. Agrawal, R. Srikant, et al., "Fast algorithms for mining association rules," in Proc. 20th Int. Conf. Very Large Data Bases, 1994, pp. 487–499.

[7] H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovery of frequent episodes in event sequences," Data Mining Knowl. Discovery, vol. 1, no. 3, pp. 259–289, 1997.

[8] K. Y. Huang and C. H. Chang, "Efficient mining of frequent episodes from complex sequences," Inf. Syst., vol. 33, no. 1, pp. 96–114, 2008.

[9] A. Achar, S. Laxman, and P. Sastry, "A unified view of the priority based algorithms for frequent episode discovery," Knowl. Inf.Syst., vol. 31, no. 2, pp. 223–250, 2012.

[10] A. Achar, A. Ibrahim, and P. Sastry, "Pattern-growth based frequent serial episode discovery," Data Knowl. Eng., vol. 87, pp. 91–108, 2013.

[11] P. Fournier-Viger, J. C. W. Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, "A survey of sequential pattern mining," Data Sci. Pattern Recognit., vol. 1, no. 1, pp. 54–77, 2017.

[12] P. Fournier-Viger ,J. Chun-Wei Lin,T. Truong-Chi,R. Nkambou," High-Utility Pattern Mining"Studies in Big Data 51, Springer 2019

[13] Yao, H., Hamilton, H.J.:Mining itemset utilities from transaction databases. Data Knowl. Eng.59(3), 603–626 (2006)

[14] Liu, M., Qu, J.: Mining high utility itemsets without candidate generation. In: Proceedings of the 21st ACM International Conference Information and knowledge management, pp. 55–64.ACM (2012)

[15] Qu, J.-F., Liu,M., Fournier-Viger, P.: Efficient algorithms for high utility itemset mining without candidate generation. In: Fournier-Viger et al. (eds). High-Utility Pattern. Mining: Theory,Algorithms and Applications. Springer (2018) (to appear)

[16] TOPIC: Top-k High-Utility Itemset Discovering Jiahui Chen, Member, IEEE, Shicheng Wan, Wensheng Gan, Member, IEEE,Guoting Chen, and Hamido Fujita, Senior Member, IEEE

[17]Chongsheng Zhang,George Almpanidise,,Wanwan Wang, Changchang Liu,"An empirical evaluation of high utility itemset mining algorithms", Volume 101, Elsevier , 2018lnidiangchangLiu

[18] Fournier-Viger, Jerry Chun-Wei Lin, Tin Truong-Chi and Roger Nkambou P.A," Survey of High Utility Itemset Mining 2019

[19] Duong, H., Ramampiaro, H., Norvag, K., Fournier-Viger, P., Dam, T.-L.: High utility drift detection in quantitative data streams. Knowl. Based Syst. 157(1), 34–51 (2018)

[20] Shie, B.-E., Yu, P.S., Tseng, V.S.: Efficient algorithms for mining maximal high utility itemset from data streams with different models. Expert Syst. Appl. 39(17), 12947–12960 (2012)

[21] Attri Ghosal, Arunima Nandy, Amit Kumar Das, Saptarsi Goswami and Mrityunjoy Panday," J. K. Mandal and D. Bhattacharya (eds.), Emerging Technology in Modelling and Graphics, Advances in Intelligent Systems and Computing 937, Springer Nature Singapore Pte Ltd. 2020

[22] S. Kaushik, An introduction to clustering & different methods of clustering (2016, December 10). Retrieved July 5 , 2018, from https://www.analyticsvidhya.com/blog/2016/11/anintroduction-to-clustering-and-different-methods-of-clustering/

[23] Data Clustering: Theory, Algorithms, and Applications. (n.d.). Retrieved July 12, 2018, from https://epubs.siam.org/doi/abs/10.1137/1.9780898718348.ch12