The Use of Machine Learning Models in Customer Segmentation on Airline, Retail and Electricity Markets

Zixuan Hu[1, *, †], Yuxin Li[2, †], Jiachen Wang[3, †]

[1]School of Information Renmin University of China Beijing, China

[2]Bachelor of Arts University of British Columbia Vancouver, Canada

[3]Computer Science Huazhong University of Science and Technology Wuhan, China

[*]Corresponding author: hzx@ruc.edu.cn

[†]These authors contributed equally.

Keywords: Customer segmentation; customer-oriented; airline; retail; electricity

Abstract: Customer segmentation is essential for customer-oriented industries. By treating customers differently, the company can win preference with a wider range of customers. In order to accomplish the goal more precisely and efficiently, several machine learning models including K-Means algorithm and Self Organized Maps are used. Customer segmentation is applied in numerous industries, among them, airline, retail market and electricity stand a dominate position, as in these areas, customer segmentation plays a more important part in making market strategies compared with other areas. Thus, this paper mainly focuses on categorical skills in these three areas, makes a timely review of the customers' classification, and puts forward the potential directions for the future. The review on customer segmentation may benefit investors for in-depth studies.

1. Introduction

Machine learning is commonly used in business analysis, as it can easily achieve some goals that is difficult or time-consuming for human. In business analysis, customers are resources for companies. Improving customer relationships and attracting more customers can somehow enhance their competitiveness in the market, thus the company can make more profit.

Customer classification has always been a key part of the data analysis and strategy making of companies. The more organized and specialized customers are classified, the more valid information the companies can get. Only when enterprises are given all the sorted information of existing customers and possible future clients can they make their own decisions accordingly. When companies plan to allocate their resources to develop new customers, how can they predict which customers will be their potential clients and which ones are unlikely to be their clients? And what proportion of the resources should be allocated to future clients, what proportion should be left to present clients? These puzzles can all be settled through customer classification by predicting the past-based future.

Customer segmentation is commonly used in multiple areas; however, most researchers concentrate on several areas and lack systematic reports on others. One of the areas that contribute a lot to the research of customer classification is the banking system, which is a competitive industry and is customer-oriented. To stand out among banks, each bank must classify customers carefully and recommend a suitable project for profits. A common classification index in the banking system is whether the customer is a risk-lover.

However, areas such as retail, electricity and airline markets lack summaries of customer segmentation and deserve in-depth studies. To the best of our knowledge, this review is of particular significance to the airline, retail, and electricity markets and makes the following contributions to the literature. First, for the airline, the great changes in the field of airline and other factors such as the Internet and policy motivated the optimization of the original customer segmentation method. In chronological order, the paper summarizes the changes of methods in this field, mainly focused on machine learning methods and several new segmentation criteria; Second, as for the retail market, this paper concentrates on the factors to be used in the classification and the related adaptation among them; Third, when it refers to the electricity market, the paper divides the whole picture into several aspects, i.e., electricity infrastructure, etc., and briefly states the different and useful algorithms. Finally, we point out the potential research direction that may benefit the subsequent investors in customer segmentation.

This paper is constructed as follows. Section 2 shows a brief review of the three representative fields, and Section 3 concludes the paper.

2. Literature Review

Customer classification is necessary for companies. It is commonly accepted that to allocate the limited resources effectively. Companies should categorize customers and find out valuable customers. In this way, companies can develop more targeted plans to serve profitable customers and avoid wasting resources on customers who are not so valuable. In addition to that, customer segmentation is extremely helpful for product selling and product development. And in the market, firms must own innovative marketing strategies and products, which make them survive.

3. Airline

With the great changes, i.e., Internet and policy, it is far from enough to serve customers by the original customer segmentation method in the airline. The classical customer segmentation divides passengers into the business type and economic type, and then air companies provide goods and services separately with flexibility for business passengers and lower prices for economic passengers. However, Teichert [1] pointed out that as the government no longer controls the airline industry, the competition among companies increases dramatically. The rise of low-cost airlines attracts not only economic passengers but also business passengers, which means that the preference differences between economy and business choices are reduced. Evidence shows that the internal competition of airlines is fierce, but other transportation industries also impact the aviation industry. For example, the high-speed rail [2]. Also, the convenience of the Internet makes the ticket price transparent, which exacerbates the changes in customer preferences and behaviors [3].

To search for proper classification standards for customers in the airline industry, some research uses machine learning methods. For example, Wong and Chung [2] applied data mining methods and the C5.0 decision tree to help segment valuable passengers. According to Berry and Linoff, the data mining method contributes to mining-related technology and algorithm, it is good at searching hidden knowledge [4]. C5.0 decision tree is an upgrade of C4.5 decision tree, and it is useful for exploring characteristics about demographic, travel actions, and service quality. Similarly, Tirenni [5] and Kim [6] also agreed that the practicability of the decision tree is fast and flexible, especially compares to neural networks. Tahanisaz [7] developed a clustering model based on the ICF model, which is similar to RFM but pays more attention to customers' common expectations. Segmenting different passengers is based on their expectations, therefore airline companies can concentrate on passengers who have high expectations in some aspects. Wei [8] also proposed a clustering method called the Ant Colony Clustering algorithm, which improves the similarity between the groups and helps wipe off strange data such as outliers. “The ant colony clustering algorithm mainly used adaptively adjusted group similarity to perform clustering and access to initial clustering result. Then all data representation points and abnormal data were inputted into a lattice plane scattered randomly. Ant colony algorithm was used for clustering once again and the corresponding class label was used to delete abnormal values and obtain complete clusters. Data test example based on ant colony clustering customer analysis platform illustrated its feasibility and effectiveness.’’Ustebay [9] and Kaski [10] applied Self Organized maps (SOM) and the K-Means method for clustering, and both of them are unsupervised learning methods. Especially, Kaski also stated that the Self-organizing maps are a machine learning method that displays the data in the form of a map, it is suitable when many fields exist, which means high dimensional. Thus, the self-organizing map (SOM) algorithm of can be used to aid the exploration: the structures in the data sets can be illustrated on special map displays. In this work, the methodology of using SOMs for exploratory data analysis or data mining is reviewed and developed further. The properties of the maps are compared with the properties of related methods intended for visualizing high dimensional multivariate data sets. According to MacQueen [11], the K-Means algorithm assigns an object to the cluster with the nearest centroid.

4. Retail market

Compared with other areas, the retail market is an especially dynamic one. This is traditional because of the similarity in the offered products since all retailers have access to the same range of products via their distributors [12]. Recently, the Internet brought out some new business concepts and further intensified internationalization and increased competitive pressure, which means that in order to match with the change in the whole market, some optimization or improvement requires to be getting into researches, concentrating on a variety of factors to be used in customer classification.

Every business organization has a primitive goal for profits. To achieve the goal, they apply sales promotion strategies to show their product and their promotion activities, such as a discount on a particular item to customers. Since, if they apply their effort in a particular direction then the intensity of effort will increase. Nowadays, there is an increasing awareness that effective customer relationship management can increase profits by understanding the needs and preferences of the customers. As a result, the problems arising in category management can be separated into four different areas [13], first, campaign optimization, i.e., selection of target groups and customers; second, cross- and up-selling, i.e., additional sales to customers; third, assortment optimization, i.e., product assortment and categories; forth, price optimization, i.e., optimization of product prices and promotions.

Since, when we decide how to classify customers in the retail market, we must consider these four kinds of problems. RFM has been widely applied for customer value analysis to solve these puzzles. Over the past twenty years, several researchers have considered RFM models in developing customer prediction and classification models. Many scholars have used it to deal with customer segmentation [12, 14-17]. Since RFM analyzes the behavior of the customers, it can be possible to encounter behavior-based models in the literature. And they can be used to classify customers in terms of their profitability and create a customer lifetime value. Additionally, using RFM variables to estimate customer’s responses is also a good choice. Loyalty program is also a commonly used customer segmentation method and has achieved numerous achievements. For example, Leenheer [18] defines a loyalty program as "an integrated system of marketing actions". A customer must become a member and identify himself as such with his loyalty card at every purchase occasion, to take advantage of the loyalty program”. The major reasons of companies’ loyalty programs usage are increasing customer loyalty, collecting customer and shopping habit data, retaining customers and selling them more rewarding frequent shoppers, and promoting customized offers. Segmenting customers and implementing more successful loyalty programs has become more easy and useful in recent years thanks to advanced data mining techniques. Kandampully and Suhartando [19] and Bulut [20] also refer to customers' repurchase behavior and the frequency of repurchases as a component of customer loyalty. The existing marketing literature is also full up of studies that reveal the relationship between loyalty program membership and brand loyalty and show remarkable results. Similarly, Ramaseshan [21] determined that demotion on ownership of loyalty programs has stronger negative effects on customers' attitudes and behavioral intentions. Customers who are members of the loyalty program show higher behavioral and attitudinal loyalty, visit retailers more than non-members, and purchase more [22]. Companies have more useful data to better segment their customers by using these techniques depending upon their competencies in data analysis and interpretation.

5. Electricity Market

It is of great significance to put customers into different categories for electricity companies, one vital reason is that customers vary in their consumption ability and willingness to pay, which leads to fluctuations in profits [23]. In addition to that, the necessity of customer segmentation is embodied more importantly in cutting down the waste of energy. Since global warming is developing rapidly, electricity companies need to be aware of the energy consumption mode of each household in order to maintain a healthy and sustainable way of development. By putting customers into different categories, companies can recommend different electricity usage plans to different customers.

Machine learning is a useful tool in the electricity market since it helps to categorize customers in various application areas, to name a few, electricity marketing, electricity tariff, electricity infrastructure building, and solving the problem of electricity theft.

When it comes to the area of marketing, machine learning methods help the industry better acknowledge the characteristics of customers, summarize users' behaviors and consumption patterns, and then formulate market strategies accordingly. In a free and comparative electricity market, the knowledge about how and when consumers use electricity plays an essential role. To utilize the dynamic information, the use of this data must be made with the application of Data Mining and Knowledge Discovery techniques to support the development of generic load profiles to each consumer's class. One way of dealing with this problem is to propose a KDD project applied to electricity consumption data from utility clients database, where it forms the different customers' classes a comparative analysis of the performance of the Kohonen Self Organized Maps (SOM) and K-means algorithm for clustering is presented [24]. Another dealing method aiming to help Electric-Power Industry to fast recognize customer's features is to propose a multi-model of customer classification. Machine learning algorithm and its strategy selection model are investigated to achieve a multi-model, which can effectively draw customer portraits and recognize targeted customers from big data [25]. Besides, the data of power utilities can also provide insights for planning outages, making network investment decisions, predicting future load growth, and predictive maintenance. Through two well-known cluster evaluation metrics, different similarity measures used in the k-means clustering algorithm are compared and analyzed, and ultimately form the most precise and efficient clustering method [26].

As for electricity providers in need of formulating dedicated tariff offers, a key aspect for structure building is the identification of the consumption patterns of the customers, in order to form specific customer classes containing customers exhibiting similar patterns. With those patterns mentioned above, a lot of things such as examining the best possible scheme of charging can be done. One solution is to compare the results obtained by using various unsupervised clustering algorithms (modified follow-the-leader, hierarchical clustering, K-means, fuzzy K-means) and the self-organizing maps to group together customers with similar electrical behavior, followed with the test of comparing their effectiveness through a set of clustering validity indicators [27]. Another way to develop and apply the patterns is to build a simulation model that is used to assess the effectiveness of demand response strategies under different time-of-use electricity tariffs in conjunction with zone thermal control [28]. With the model precisely set up on the ground of data, it will be very helpful in further studies.

When building electricity infrastructure, smart grid technologies and wide-spread installation of advanced metering infrastructure equipment have presented new opportunities for the use of machine learning algorithms paired with big data to improve distribution system models. Accurate models are critical in the continuing integration of distributed energy resources into the power grid. To form such a model, we need to propose a novel spectral clustering approach for validating and correcting customer electrical phase labels in existing utility models. Spectral clustering containing creating affinity matrix, nonlinear dimensionality reduction, and K-Means clustering is used to improve the accuracy and scalability of the algorithm for large datasets [29]. Also, it is very crucial to update the customer feedback system along with the electricity infrastructure, since knowing the percentile of their consumption among similar families helps them to think of ways to reduce their own electricity usage, which decreases the emission of greenhouse gases. The feedback system can be improved by firstly segmenting the customers by the heating system and the type of housing, followed by weighted clustering used to refine the comparison group, and eventually, understandable customer-specific comparison information and easy-to-use energy displays can be presented [30].

6. Conclusion

In the whole market, it is the most significant for firms to acquire innovative marketing strategies and products, which is essential for them to survive and thrive. Under this condition, appropriate customer classification methods are able to help satisfy customers, which can, to a great extent, devote to cultivating diehard fans and attracting new customers. Consequently, those customers can then make a great contribution to the interest of the company. As representatives of industries that apply customer segmentation methods to develop companies' strategies, the airline, retail market, and electricity industry show how much benefits it can bring through utilizing these methods when encountering some appropriate occasions. Putting into consideration what has been mentioned above, this paper overall makes a review around customers classification focusing on these three areas. In short, when it refers to the airline, machine learning is the technique with the highest efficiency, particularly the K-Means algorithm, and this paper also concludes some special but meaningful factors in order to draw a more detailed and advanced picture. As for retail markets, people always solve this area of problems depending on four parts and adopt diverse parameters and weights aiming at different issues. Furthermore, electricity marketing often uses machine learning methods to assist the industry in better acknowledging customers' characteristics and accordingly arrange to formulate market strategies. The review on customer classification may benefit related participants in academia and industry. However, through the review, it can also be noted that methods, i.e., machine learning, K-Means, etc., are widely applied in customer classification. Adopting alternative methods to classify customers may deserve further investigation.

References

[1] Teichert, T., Shehu, E., von Wartburg, I.(2008). Customer segmentation revisited: The case of the airline industry. Transportation Research Part A: Policy and Practice, 42(1): 227-242.

[2] Wong, J. Y., Chung, P. H. (2007). Managing valuable Taiwanese airline passengers using knowledge discovery in database techniques. Journal of Air Transport Management, 13(6): 362-370.

[3] Lindstädt, H., Fauser, B. (2004). Separation or integration? Can network carriers create distinct business streams on one integrated production platform?. Journal of Air Transport Management, 10(1): 23-31.

[4] Berry, M. J. A., Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.

[5] Tirenni, G., Kaiser, C., Herrmann, A. (2007). Applying decision trees for value-based customer relations management: Predicting airline customers' future values. Journal of Database Marketing & Customer Strategy Management, 14(2): 130-142.

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

[7] Tahanisaz, S. (2020). Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry. Journal of Air Transport Management, 83: 101764.

[8] Wei, L. F. (2012). Design and Implementation of Airline Customer Segmentation System Based on Ant Colony Clustering Algorithm. Advanced Materials Research. Trans Tech Publications Ltd, 433: 3357-3361.

[9] Ustebay, S., Yelmen, İ., Zontul, M. (2020). Customer Segmentation Based on Self-Organizing Maps: A Case Study on Airline Passengers. Journal of Aeronautics & Space Technologies/Havacilik ve Uzay Teknolojileri Dergisi, 13(2).

[10] Kaski, S. (1997). Data exploration using self-organizing maps.Acta polytechnica scandinavica: Mathematics, computing and management in engineering series no. 82.

[11] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1(14): 281-297.

[12] Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. Journal of Retailing and Consumer Services, 61, 102566

[13] Garcke, J., Griebel, M., Thess, M. (2010) Data Mining for the Category Management in the Retail Market. In: Grötschel M., Lucas K., Mehrmann V. (eds) Production Factor Mathematics. Springer, Berlin, Heidelberg

[14] Spring, P., Leeflang, P. S. & Wansbeek, T. (1999). The combination strategy to optimal target selection and offer segmentation in direct mail. Journal of Market-Focused Management, 4(3), 187-203

[15] Jonker, J. J., Piersma, N. & Van den Poel, D. (2004). Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. Expert Systems with Applications, 27(2), 159-168

[16] Cheng, C. H. & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications, 36(3), 4176-4184

[17] Khajvand, M. & Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. Procedia Computer Science, 3, 1327-1332

[18] Leenheer, J. & Bijmolt, T. H. A. (2008). Which retailers adopt a loyalty program?An empirical study. Journal of Retailing and Consumer Services, 15(6), 429-442

[19] Kandampully, J. & Suhartanto, D. (2000). Customer loyalty in the hotel industry: The role of customer satisfaction and image. International Journal of Contemporary Hospitality Management, 12, 346-351

[20] Bulut, Z. A. (2015). Determinants of repurchase intention in online shopping: a Turkish consumers' perspective. International Journal of Business and Social Science, 6(10), 55-63

[21] Ramaseshan, B., Stein, A. & Rabbanee, F. K. (2016). Status demotion in hierarchical loyalty programs: Effects of payment source. The Service Industries Journal, 36(9), 375-395

[22] Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. Journal of Marketing, 71(4), 19–35

[23] Dutta, G., & Mitra, K. (2017). A literature review on dynamic pricing of electricity. Journal of the Operational Research Society, 68(10), 1131-1145.

[24] Figueiredo, V., Duarte, F. J., Rodrigues, F., Vale, Z., & Gouveia, J. (2003, September). Electric energy customer characterization by clustering. In Proc. ISAP (pp. 1-6).

[25] Ruan, Q. S., Wu, Q. F., Tseng, H. W., & Liu, X. L. (2017, November). A Multi-Model of Classification for Electric-Power Industrial Customer Based on Big Data. In 2017 International Conference on Information, Communication and Engineering (ICICE) (pp. 562-564). IEEE.

[26] Du Toit, J., Davimes, R., Mohamed, A., Patel, K., & Nye, J. M. (2016). Customer segmentation using unsupervised learning on daily energy load profiles. J Adv Inform Technol, 7(2).

[27] Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification. IEEE Transactions on power systems, 21(2), 933-940.

[28] Pallonetto, F., De Rosa, M., Milano, F., & Finn, D. P. (2019). Demand response algorithms for smart-grid ready residential buildings using machine learning models. Applied energy, 239, 1265-1282.

[29] Blakely, L., Reno, M. J., & Feng, W. C. (2019, February). Spectral clustering for customer phase identification using AMI voltage timeseries. In 2019 IEEE Power and Energy Conference at Illinois (PECI) (pp. 1-7). IEEE.

[30] Mononen, M., Saarenpää, J., Johansson, M., & Niska, H. (2014, April). Data-driven method for providing feedback to households on electricity consumption. In 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) (pp. 1-6). IEEE.