AN ADVANCED CLUSTERING MODEL FOR MOBILE CUSTOMER RELATIONSHIP MANAGEMENT SYSTEM

Mr. D. MADHAVA RAO[1], Mrs. V. TEJASWINI[2*]

1 Final Year MCA Student, QIS College of Engineering and Technology, Ongole

2* Assistant Professor, MCA Dept., QIS College of Engineering and Technology, Ongole

Abstract: *Customer segmentation provides an efficient way to acquire the insights into client characteristics and activity preferences. So as to reinforce client relationship management and to specialise in a dynamically changing market place, we've engineered a two-layer model for mobile telecommunication client analysis. The primary layer identifies the cluster by domain knowledge in accordance with the contribution of the essential degree of client segmentation. The second layer uses clustering algorithms to cluster along extremely consistent teams and determine representative teams so as to clarify the characteristics of the cluster. It then builds a mobile-user cluster model supported client attributes, contribution to assist enterprises read changes in client's behavior. The clustering results of this study are legible, have marketing implications, and are intuitive. They'll be combined with the merchandise to help employees with effective marketing. After verification, this methodology will attain the specified results.*

Keywords**: **Data mining, Data analysis, Cluster segmentation, Business intelligence.

I. INTRODUCTION

With the development of big data & data mining technology, the mass storage of internal enterprise data are often analyzed effectively with the hidden client value that may be found by mining. Promotion activities and relationship support is additionally supported a antecedently in depth marketing model that was progressed to focus on the client base, acquire in-depth understanding, and work the requirements of the client base of precision marketing. An increasingly important issue is how to integrate the marketing resources, and how to properly distribute and match individual customer interests and preferences with the best benefit from marketing activities, as well as mining attractive products or services [1].

Data clustering and clustering algorithms provide an efficient way to group highly homogeneous individuals and assign dissimilar individuals to the appropriate segments [2]. In trade and world, there are many examples of clustering analysis being used to establish cluster characteristics: "customer grouping analysis" is a popular application

By analyzing client attributes, behaviors, and preferences, the high homogeneity of individual clusters, and also the high degree [3], of unsimilarity of people given the acceptable phase. additionally to effectively manage the homogeneity of the characteristics of high client base, may be used as enterprises for the restricted client base to develop management ways and selling principles of the muse to help client relationship management.

During this study, we use the activity knowledge of mobile users to ascertain a mobile-user agglomeration model by analyzing client attributes, client contributions [4], and cluster segmentation. We propose a two-layer agglomeration model to assist enterprises read changes in customer vaue and behavior. The two-layer agglomeration model projected during this study may be applied to alternative business areas that may track customer behavior, like the membership cards that are employed in retail sales (e.g., COSTCO) or the provision of credit cards by banks. Through such cards, client consumption are often recorded and known, and also the projected two-layer agglomeration model is often used for business analysis.

The rest of this paper is organized as follows. Section II provides an outline of client agglomeration studies. Section III describes the two-layer model of mobile-customer agglomeration and segmentation strategies. Section IV illustrates the characteristics of agglomeration, client interpretation, and also the results of the analysis. Finally, conclusions and future work area unit conferred in Section V

II. Introduction to Customer Clustering

Clustering algorithms will be divided into the subsequent general categories: 1) hierarchical 2) partitional 3) density-oriented 4) grid oriented 5) model-based—existing models

Cluster analysis is widely used for variable knowledge analysis in fields like medication, economics, text mining, and industrial applications. There are several studies on cluster analysis for separating knowledge characteristics and detective work knowledge clump phenomena. There square measure several business applications, like target/direct promoting supported client grouping and customization services, smart customer-relationship management, more over as client behavior, attributes, and preferences.

In keeping with the 80/20 rule (or Pareto's Principle) [5], eightieth of a company's profits come back from the foremost vital two hundredth of its customers, with the remaining two hundredth of profits returning from the standard eightieth of customers. If an organization will absolutely comprehend its key two hundredth of customers, those customers will bring a considerable profit to the corporate. The connected analysis combines the ideas of client time period worth (CLV) and client segmentation. Customers type applicable segments to assist the corporate target its target customers (TA) so develop customer-relationship management, promoting ways, and promotional activities.

Client time period worth refers to the entire revenue that every client will arouse the enterprise. It will be divided into the customer's historical worth, current worth, and potential worth. Tutorial analysis on client grouping has been conducted supported one hundred fifty-five, and 3 completely different models are advocate in keeping with client contribution, basic attributes (e.g., age and gender), and most popular client behavior [6]. The results show that client teams fashioned through multiple dimensions will differentiate client attributes effectively. Cutting the bulk of the client teams into variety of special behavioural subgroups helps the corporate gain Associate in nursing in-depth understanding of its client base [7].

However, most client grouping so far has either been supported rules of thumb or has paid attention solely to the typical revenue per user (ARPU) as a benchmark for client segmentation. Solely a number of approaches have enclosed alternative factors like client life cycle or overall client contribution [8, 9]. Vodafone, a British telecoms operator, segments mobile users into several solid clusters through "customer segmentation" and "customer profiling" to spot common options. It uses the analysis and outline of user attributes to assist with management decision-making and operational pointers [10].

During this study, we propose a two-layer clump model supported the analysis of client attributes, client contributions, and cluster segmentation. We cluster the worth of mobile customers and execute the characteristics of customers [11, on a daily basis in an exceedingly systematic manner. Preference analysis will facilitate an organization read changes in client worth and behavior, and at any time to regulate its product strategy to keep up high-quality customers. Our model provides how for firms to arrange for long customer-relationship management and maintain high quality customers. Additionally, sellers will use this modeling [12], approach to market product or services accurately.

III. Two-layer Clustering Model

A telecoms business can have as many as 10 million mobile customers. If there is no further customer segmentation in accordance with the business, characteristics and other conditions to distinguish. Not only is it difficult to grasp the dynamic changes in customer management, the retention for customers will also cause huge maintenance costs. Therefore, this study will provide telecommunications companies based on operational needs (including mobile, data, and other business) in accordance with the two-layer clustering model for customer segmentation. Figure 1 shows a flowchart of the two-layer clustering model.

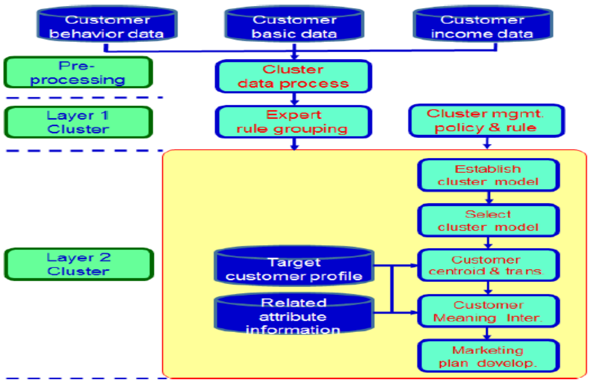

Fig. 1. Flowchart of the two-layer clustering model

This method begins by collecting and segmenting the contribution of individual customers, their personal preferences, overall customer profile, and other factors. The overall customer segmentation for a number of customers is in the order of 105–106 clusters. These clusters are then used to develop a general strategy for customer-relationship management, which forms the first layer of clusters. After the first layer of the target has been clustered, the characteristics of the subgroups are described and interpreted by the subdivision of the secondlayer clustering algorithm. With the big data platform data cross-analysis function, to further develop the maintenance of each group of customers, as well as effective marketing programs to form the second layer of customer clustering analysis.

The aims of the proposed two-layer clustering model are as follows:

1) To provide real-time, diverse, and rich customer information through pre-planned, pre-analysis to strengthen the target customer base and reduce the workload of marketing staff;
2) To evaluate the customer segmentation strategy for each group to improve the effectiveness of activity planning and the customer-relationship management strategy;
3) To use data mining technology to tap potential target customers, increase the feasibility of marketing products and services, and improve the accuracy of precision marketing.

In the model proposed in this paper, the first layer inspects the customer value, and the second layer uses consumer-behavior features to further group. In practical applications, the definition of customer value varies by industry. Even the same industry can have different priorities, such as the amount of consumption, the number of consumers, the number of stores, and so on. Therefore, when implementing this model, each industry must first define the customer value of each variable. In addition, because customer behavior may change, in order to make the marketing strategy more accurate, we must dynamically monitor the changes in these acts, coupled with automatic mechanisms for observing customer-behavior trends and providing early warnings. The proposed twolayer clustering model of the segmentation approach is described below.

A. Layer 1 Clustering Architecture

The first-layer cluster architecture (hereinafter referred to as the L1 cluster) uses mean and divides the customers into two groups according to two-dimensional attributes, as shown in Fig. 2. The horizontal axis divides the overall mobile-customer base in the range 0–99 (i.e., a total of 100 rankings) according to the customer's contribution to the company's revenue. The higher the value, the higher the contribution on behalf of the customer, and the higher the customer's value to the company. The vertical axis is based on voice-leased monthly bills. The higher the voice-call monthly fee, the greater the customer's reliance on the telecom's mobile service, and the higher the demand for mobile calls. Finally, the behavior of special data-oriented billed users (customers whose main need is for mobile Internet) becomes an independent group, S7. In this study, we take the contribution and ARPU as the customer's grouping variables. In practical applications, the definition of customer value varies according to each industry. Even if the same industry may have different priorities, at this time, through the small scale test to determine the first layer of the selected group variable.

Mobile customers in the first layer can be divided into S1–S7, a total of seven large groups. A separate overall strategy for customer-relationship management is developed for each group. For example, in accordance with the L1 cluster cutting, the overall strategy for maintaining the customer base should be driven by the performance of the S1 customer group as the main criterion.

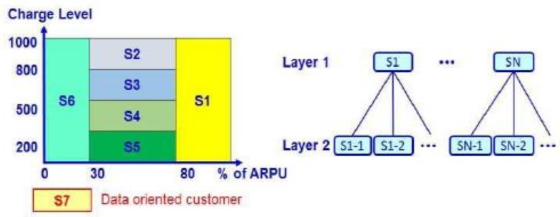

Fig. 2. Schematic of Layer-1 and Layer-2 clustering

Among the customer groups, S1 makes the highest contribution to the customer base. Focus should be committed to maintaining that existing customer base and strengthening customer loyalty in order to stabilize the company's revenue. Meanwhile, the S6 group makes a low contribution to the customer base. The focus here should be on enhancing customer value and strengthening the dependence on mobile services in order to enhance customer contribution to the company's revenue. In addition to the special behavior of the S7 customer base, the company should strengthen the promotion of its value-added services to drive customer demand for voice services.

B. Layer 2 Clustering Architecture

The second-level grouping (hereinafter referred to as the L2 cluster) is structured under the L1 cluster of expert rules. Subgroups are subdivided for each L1 cluster, as shown in Fig. 2. First of all, we aggregate for each L1 cluster by calculating the customer's communication-use behavior to distinguish between the behavior of customer preferences and their business habits. The cluster variables used by the L2 cluster subdivision may vary depending on whether the customer belongs to different L1 clusters.

Taking mobile-customer segmentation as an example, in general postpaid customers S1–S6 group. In order to find out the main sources of communication behavior and customer contribution, thus adding the inter/intra network traffic minutes, the number of called objects, the ratio of each sub-item to the total bill and other variables.

In contrast, because the large S7 group is of customers whose main need is for mobile Internet (i.e., a dataoriented type), its call behavior and use of other mobile value-added services are significantly different from those of general monthly post-paid customers. Variables assigned to the second-layer cluster will focus on the use of mobile data. For example, the proportion of downloads, the growth rate of data transmission, and the time proportion of data transmission can be used to outline the behavior of customers' online-behavior variables rather than voice-behavior variables.

In order to make the appropriate segmentation among customers in S1–S7, this study adds the customer's ranking in the grouping of customers in the group variable design. For example, international voice calls are telecommunications services that are often used by international roaming or business people both at home and abroad, and are a source of mobile revenue. To differentiate the main users of international voice services into different cost bands and customer contributions. This paper investigates which customers are willing to use international voice roaming services and establishes the difference in the dependence degree of international voice roaming groups. We summarize the number of minutes used by individual international voice customers in the overall rankings of the customer segments in international phonetic variables to derive the variable type. Therefore, after the second-level grouping, the relative subgroups in the overall group international voice call behavior can analysis.

In this study, the following steps are used to clean up (pre-process) the input variables of the L2 cluster:

1) Correlation coefficient analysis—eliminate the dependency on high-value variables in order to avoid excessively similar weightings that result in sub-group errors;
2) Outlier processing—replace the outliers using the ceiling/floor method to avoid bias due to extremes that affect the clustering algorithm in determining customer attributes.

After determining the two-layer clustering model and segmenting the L1 cluster into clustering variables, this study uses SPSS Clementine 13.0 to implement the L2 clustering. Clustering algorithms such as *k*-means, twostep, and Kohonen are used to establish the cluster model according to the likes of customer attributes, behavior preferences, and contract status. Each L1 cluster is split into between five and seven different L2 clusters. The optimal clustering model is then selected as the basis of the final L2 clustering, based on conditions such as the maximum and minimum cluster ratio, silhouette coefficient, and marketers' readability.

IV. Results of Clustering Analysis

Currently, many indicators are used to evaluate the clustering results [11, 12]. In general, a good clustering method should match the criteria that the clusters are with similar size, the distance of an object to other neighbor cluster is long, but the distance between objects within the same cluster is short. In this study, the following criteria are used to select the optimal cluster in the L2 grouping.

(1) Cluster number and cluster size

The number of L2 clusters and the number of customers that belong to each cluster should be based on the range of activities that the company's marketers can promote. It should avoid over-segmentation of customers, and clusters that are either too large or too small.

(2) Value of silhouette coefficient

The silhouette coefficient is a measure of the distance between groups, as shown in Fig. 3. It is combined with cohesion and dispersion to determine the quality of the sub-group of indicators. Generally, the silhouette coefficient is greater than 0.2; if it is greater than 0.5, the cluster is more effective in distinguishing between individual and heterogeneous/ homogeneous individuals.

(3) Maximum and minimum cluster ratio

The maximum–minimum cluster ratio represents the ratio of the maximum number of clusters to the number of individuals in the smallest cluster. A lower maximum–minimum cluster ratio means that the size of each sub-group is closer to the average, to avoid all individuals being concentrated in a large group of bias. This study used a lower ratio.

(4) Cluster readability

The characteristics of a cluster should make marketing sense, be intuitive, and be aligned with the company's products or services to allow marketing staff to market directly and effectively.

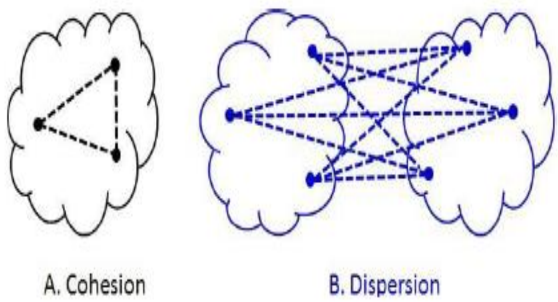

Fig. 3. Cohesion and dispersion of groups

There follows a discussion of the cluster interpretation and the development of marketing policy. In order to gain a deeper understanding of cluster characteristics, the cluster center (centroid) of each L2 cluster is output after grouping to provide a marketing policy for defining the sub-groups that are actually available for operation. The cluster center refers to the representative point of each L2 cluster, and can be expressed simply as the average value of each variable of the group. It can explain the distribution of the L2 cluster variables. In order to facilitate the interpretation of the values, the distribution of the cluster centers of all L2 clusters is converted into five classes according to the mean value (μ) and the standard deviation (σ) of the parent variables (see Fig. 4). This assists in comparing each L2 cluster for differences in overall customer characteristics. In addition, the integration of the basic information of mobile customers (e.g., contract status, age, and gender) and other cross-analysis results gives a basis for the interpretation of cluster characteristics.

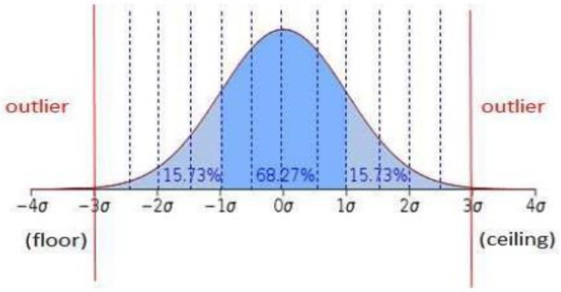

Fig. 4. Class center-level distance conversion diagram

Table 1 gives an example of cluster center transformation and feature description for L2 mobile clusters. It lists the cluster center of each subgroup and the cluster center distance after conversion. For example, μ = 7.20 for the total customer and σ = 47.14 for the global variable, which is the ratio of the international call charge to the bill amount. Therefore, the cluster center of S1-3 is 84.76, which is greater than 54.34 (i.e., μ + σ). The center of the group is converted to "high" at the level at which the international telephone bill occupies the bill amount ratio according to the classification of the group center pole pitch conversion.

Table1. Cluster center transformation and description for L2 mobile clusters

Variable Name	S1-1		S1-2		S1-3		S1-4		S1-5	
Monthly charge / Amount %	376.65	ML	427.98	ML	431.85	ML	539.35	M	599.65	M
International bill / Amount %	1.20	M	27.35	M	84.76	H	1.04	M	0.80	M
Data bill / Amount %	516.71	H	518.73	H	162.93	M	36.45	M	13.07	M
VAS bill	23.18	M	1192.66	H	45.32	M	10.06	M	6.04	M
Ranking of amount charge	902.00	H	829.00	H	910.00	H	868.00	H	694.00	MH
Proportion of inter. Net calls	434.37	M	446.58	M	431.87	M	438.84	M	526.51	M
Total number of calling objects	22.12	M	29.38	MH	42.23	H	47.95	H	19.83	M
Ranking of local call minutes	580.00	M	645.00	MH	778.00	H	853.00	H	404.00	M
Ranking of inter. call minutes	0.00	N	64.00	M	967.00	H	0.00	N	2.00	M

Remarks: H = high, M = middle, L = low, N = null

After the group center is converted to a class distance, it can be used to describe the characteristics of each group and formulate a marketing strategy. For example, S1 is a large group of L1, and can be divided into five L2 clusters, such as S1-1 to S1-5. However, there are differences in the nature of the subgroups, such as the S1-3 cluster in which the ratio of international calls to invoices is significantly higher than for the other four subgroups. To get a more accurate target client, L2 selects the appropriate algorithm, such as 2-steps, K-means, and silhouette coefficient.

Moreover, with the basic cross-analysis of customer data, S1-3's use of international roaming accounted for 13.1%, customers aged 30–45 accounted for slightly more than all customers in this sub-group, and telephone rental for more than 10 years accounted for up to 40% of the sub-group or even 53.18% for customers with a smart phone. From this, we can conclude that the customer base is likely to be business travelers, for whom voice calls, international roaming concessions, and other related products are suitable.

In addition to S1-4 of the customer base, compare it in the distribution of variables and the overall customer base no significant difference. However, the rank transitions of these three items are displayed as "high" in the rank of the total bill amount, the total number of uttered objects, and the telephone voice talk time. This shows that the S1-4 target group comprises a wide range of people, suggesting that the group may serve as a sales or customer service, such as large-scale use of telephone contact with the customer base. It is recommended for a large number of calls minutes, with a discount voice package.

In order to verify the correctness of the model, approximately 7.7 million mobile consumers were clustered in the above case. The original data were stored in an SQL Server and Clementine 13.0 was used as the implementation tool for L2 clustering. We use the correlation coefficient matrix to remove variables with high correlation coefficients, and the analysis of variance (ANOVA) for cluster and input variables. With S1-3 of about 160,000 customers, marketing staff based their sub-group characteristics on the CRM development of the corresponding policy, including recommended products containing 1) a new smartphone and 2) international dataroaming day-rental products. In addition to maintaining the customer's plan includes: 1) proactive marketing of new mobile phones and international telephone promotion rates, 2) proactive notification of important customers about new smartphone sales, and 3) proactive notification of key customers about international voice and dataroaming promotions.

V. CONCLUSIONS AND FUTURE WORK

In this paper, a two-layer customer clustering model is proposed, providing a macro and micro perspective to assist mobile-customer-relationship management. Marketers can use the technology of pre-analysis and data mining to target their customers and sell the company's products and services with accurate marketing. In addition, the expert-rule L1 subgroup can also help companies to develop their general direction of customer relationship management to enhance the purpose of customer service quality.

After the clustering model and related strategies have been established, we can track the changes of the group structure periodically and systematically. This allows us to monitor trends in group movement, monitor the size change of each cluster, and adjust the group marketing policy and management strategy to achieve customer status change the effectiveness of early warning mechanism.

At present, customer clustering is only included in the cluster modeling through mobile voice, data usage behavior, customer contribution, and customer base data. In future work, we intend to increase the grouping of the customer-variables selection function. For different marketing or business needs, a customer-clustering model will be established to increase the flexibility of customer-clustering applications. In addition, in accordance with the customer group structure changes to achieve the threshold set value, the establishment of re-start the cluster modeling process or modify the marketing strategy of the warning mechanism to improve the dynamic feedback model grouping benefits.

VII. REFERENCES

[1] Ning Lu, et al., "A Customer Churn Prediction Model in Telecom Industry Using Boosting," *IEEE Trans. on Industrial Informatics*, vol. 10, no. 2, May 2014

[2] T. Jiang and A. Tuzhilin, "Improving Personalization Solutions through Optimal Segmentation of Customer Bases," *IEEE Trans. On Knowledge and Data Engineering*, vol. 21, no. 3, Mar. 2009.

[3] C. Qiuru, et al., "Telecom customer segmentation based on cluster analysis," 2012 Int. Conf. on Computer Science and Information Processing (CSIP), 2012, pp. 1179-1182.

[4] J. Han, M. Kamber, "Data Mining: Concepts and Techniques, second edition," Elsevier, San Francisco, USA, 2006.

[5] R. S. Duboff, "Marketing to maximize profitability," *The Journal of Business Strategy*, vol. 13, no. 6, 1992, pp. 10-13.

[6] S. Y. Kim, T. S. Jung, E. H. Suh, H. S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert System with Applications*, vol. 31, 2006, pp. 101-107.

[7] S. H. Han, S. X. Lu, S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert System with Applications*, vol. 39, 2012, pp. 3964-3973.

[8] S. Masood, et al., "Customer segmentation and analysis of a mobile telecommunication company of Pakistan using two phase clustering algorithm," Int. Conf. on Digital Information Management (ICDIM), 2013, pp. 137-142.

[9] J. Wu, Z. Lin, "Research on customer segmentation model by clustering," ICEC '05 Proceedings of the 7th international conference on Electronic commerce, 2005, pp. 316-318.

[10] S. M. H. Jansen, "Customer segmentation and customer profiling for a mobile telecommunications company based on usage behavior: a Vodafone case study", 2007.

[11] D. M. Eler, et al., "Simplified Stress and Simplified Silhouette Coefficient to a Faster Quality Evaluation of Multidimensional Projection Techniques and Feature Spaces," 19th Int. Conf. on Information Visualisation, 2015, pp. 133-139.

[12] H. Wang, et al., "An approach for improving K-means algorithm on market segmentation," Int. Conf. on System Science and Engineering, 2010, pp. 368-372.

Authors Profile:

Mr. **D. Madhava Rao** pursuing MCA 3rd year in Qis College and Engineering and Technology in Department of Master of Computer Applications, Ongole.

Mrs. **V.Tejaswini** is currently working as an Assistant Professor in Department of Master of Computer Applications in QIS College of Engineering & Technology.