Research Article

Application of Customer Segmentation for Electronic Toll Collection: A Case Study

Chao Qian , **Meng Yang** , **Peiqi Li**, and **Shuguang Li**

School of Electronic and Control Engineering, Chang'an University, Middle-Section of Nan'er Huan Road, Xi'an, Shaanxi 710064, China

Correspondence should be addressed to Chao Qian; chinaqianchao@163.com

Received 24 January 2018; Accepted 14 July 2018; Published 9 August 2018

Academic Editor: Giuseppe Musolino

Copyright © 2018 Chao Qian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Applying big data technology, this study presents a customer segmentation method of Electronic Toll Collection (ETC) based on vehicle behavioral characteristics. A segmentation index system of ETC customers comprising Recency, Frequency, and Monetary is extracted and constructed using ETC data. The whole-sample clustering analysis of ETC customers is accomplished with the Clustering LARge Applications (CLARA) algorithm while overcoming the invalidation problem of big data clustering. A decision tree on ETC customer segmentation is constructed and transformed into a set of segmentation rules. Empirical results indicate that the proposed method is better able to analyze travel characteristics and to present values and appreciation potentials for ETC customer classification. This method provides an innovative idea for implementing precision marketing and establishing hierarchical discount rates for ETC customers. Furthermore, it provides theoretical support to increase the ETC customer scale and payment ratio, thus improving the decision-making level in expressway operation and management.

1. Introduction

Electronic Toll Collection (ETC) is an essential part of the Intelligent Transportation System (ITS). ETC not only reduces travel time and energy consumption but also saves infrastructure and operation costs; thus, its advanced payment system is highly praised around the world. By the end of February 2017, 29 of 31 provinces in mainland China (except Tibet and Hainan) had realized networking of expressway ETC and cumulatively built 14,285 ETC lanes, 1,115 self-supporting service centres, and 37,502 cooperative agency centres. The number of ETC customers exceeded 47.67 million, and the daily average transaction number was over 8.1 million, accounting for 31.17% of the total traffic volume [1].

Since the 1990s, along with remarkable development in customer-oriented management, Customer Relationship Management (CRM) proposed by the Gartner Group Consulting Company has attracted extensive attention worldwide [2, 3]. CRM provides reliable, comprehensive, and complete understanding for enterprises through the application of emerging technologies to integrate customer data efficiently,

helpfully maintaining and expanding a mutually beneficial relationship between customers and enterprises. Aiming to allocate service resources rationally and implement customer strategies accurately, customer segmentation classifies and evaluates types of customers, thus providing theoretical and methodological guidance for enterprises' gain of higher commercial value for customers.

Any efficient CRM needs a strong foundation of customer segmentation research. Currently, research on CRM has mainly focused on the telecom industry [4–6], the energy supply industry [7, 8], and the retail industry [9]. In the automobile dealership field, Tsai et al. (2015) considered customer transaction behavior and customer satisfaction variables, using customer segmentation to develop marketing strategies [10]. Some studies have also been conducted on the customer segmentation method in different transportation modes such as railways and aviation. For instance, Wei (2012) proposed and designed a segmentation system structure for airline customers based on ant colony clustering [11]; Teichert et al. (2008) proposed an airline customer segmentation approach by analyzing more than 5800 airline passengers' stated preference data [12]. Chiang (2017) proposed a model

TABLE 1: ETC data format.

Field	Data type	Field description
CardNo	Char(20)	ETC card number
EntryTime	Datetime	Date and time a vehicle enters an ETC lane
ExitTime	Datetime	Date and time a vehicle exits ETC lane
VehicleClass	Int	Vehicle class (passenger vehicle: 1–4)
VehicleType	Int	Vehicle type (passenger / freight vehicle)
ETCMoney	Int	Payment amount
...	...	Other fields

to discover valuable travellers for airlines and generated useful association rules to find an optimized target market for CRM systems [13]. As for railways, Cheng and Huang (2014) examined the influence of ticketing channel attributes on high-speed rail passengers' preferences and designed appropriate ticketing channel services for certain types of passengers [14]. Zhang and Peng (2017) proposed a k -means based segmentation model for railway freight customers [15]. Zhong and Guo (2008) clustered freight-customer history data and classified new customers using Bayesian classifiers [16]. Duan et al. (2016) operationalized approaches to identify market segments for rail freight services and measured the importance that customers attach to rail service attributes (i.e., transport cost, time, frequency, reliability, and safety) [17].

In highway transportation, many studies have been conducted on ETC implementation [18, 19], focusing on the analysis and evaluation of the transition phase toward this new technology, as well as cost-benefit evaluation during construction, reconstruction, or an extension period of ETC [20, 21]. Astarita et al. (2001) designed a microscopic traffic simulation model to evaluate the operational efficiency of a toll station after ETC system was progressively introduced. The results indicated that the limited capacity of manual toll gates could lead to queues spill back, interfering and reducing ETC gates capacities [22]. Zarrillo et al. (2009) emphasized the significant influence of customer satisfaction to ETC usage rate and suggested that providing an appropriate incentive for regular commuters to convert from manual usage to ETC usage would be the best way to increase the throughput of a toll plaza [23].

The current research of ETC data application mainly concentrates on traffic information extraction. Ozbay et al. (2011) demonstrated that real-time travel time could be accurately estimated using ETC data [24]. Furthermore, Yang and Ozbay (2015) illustrated the potential of ETC data mining for travel time estimation for both incident-free and incident conditions [25]. However, to date, academic literature on ETC customers is relatively rare. How to obtain consumption characteristics and tap ETC customers' payment potential by analyzing the massive ETC data, to enhance customers' value and realize precision marketing, are critical problems confronting ETC promotion and application.

This work's primary goal was to establish a customer segmentation method based on ETC consumption characteristics by applying big data analysis and mining technology. A

segmentation index system was established, ETC customers were classified into categories of one to five stars, and a set of segmentation rules were extracted. In the end, travel characteristics and service strategies for each customer type were analyzed.

2. Materials and Methods

2.1. Segmentation Index. With consumption demand as a starting point, customer segmentation divides customers into similar consumer groups according to differences in their purchasing behavior. Customers on the same base have a certain degree of similarity, but customer bases show distinctions [26]. Customer segmentation models based on Recency (R), Frequency (F), and Monetary (M) or RFM behaviors, which was proposed by Hughes (1994) [27], are widely used. In this model, R represents how recently customers purchased, F how often they purchased, and M how much they spent (each time on average). Hughes believes that R, F, and M have the same degree of importance in measuring customers, and, therefore, each receives the same weight. Meanwhile, through empirical analysis of credit card data, Stone (2007) asserts that each index's weighting in customer segmentation is not the same; F should be the highest, R second, and M the lowest [28].

Expressway ETC data records various kinds of travel information, including, for example, ETC card information, travel time, vehicle information, and consumption situation. Table 1 lists ETC data's detailed format.

Each ETC datum represents an ETC customer's consumption record on a trip. An ETC customer's annual consumption can be summarized and analyzed via data aggregation. ETC customers' segmentation indexes are defined as recent consumption interval, annual frequency, and annual consumption amount (Table 2).

Hence, each ETC customer's annual consumption was aggregated according to its card number. For a particular ETC customer, assuming a frequency of F , then the indexes of R and M are calculated as follows:

$$R = T_{set} - T_{F.out} \quad (1)$$

$$M = \sum_{i=1}^F S_i \quad (2)$$

where T_{set} represents a specified time, $T_{F.out}$ means the F th consumption time in the statistical year (driving through an

TABLE 2: Segmentation index of ETC customers.

Index	Abbreviation	Unit	Meaning
Recency	R	Hour	Time difference between the last time ETC used during the statistical year and a specified time.
Frequency	F	Times	Number of times ETC was used for a particular customer during the statistical year.
Monetary	M	Chinese yuan	How much paid for ETC for a particular customer during the statistical year.

Input: D - ETC customer index dataset; k - the number of clusters; $samples$ - number of samples to be drawn from the dataset; $sampsiz$ - number of observations in each sample.**Output:**

The clustering results of ETC customer.

Method:(1) for $i = 1$ to $samples$, repeat (a)-(d);(a) select $sampsiz$ objects randomly from ETC customer index dataset D as a sample, apply the PAM algorithm to compute the best k -medoids - $[M_1, M_2 \cdots M_k]^T$;(b) apply k -medoids to the entire dataset D and calculate the distance from every non-medoids object in D to the closest object in the set $[M_1, M_2 \cdots M_k]^T$, reassign each ETC customer to different clusters;(c) compute the average dissimilarity of this clustering, if the value is less than the current minimum value, then replace the current value, and form the best k -medoids and the new set of k representative objects;

(d) return to step (1), repeat the iterative process;

(2) until no change, output clustering results of ETC customer.

ALGORITHM 1: CLARA algorithm.

ETC exit lane), and S_i denotes the monetary value of the i th time paid for ETC.

2.2. Customer Clustering. A three-dimensional space of RFM indexes was obtained from the segmentation index system previously mentioned. ETC customer clustering analysis means grouping the index dataset in such a way that index data in the same group (called a cluster) are more similar (in one sense or another) to each other than to those in other clusters. This task can be summarized as making the distance between clusters as long as possible and minimizing distances from the same cluster, thus obtaining a classification method for multi-class ETC customers.

Partition-based clustering methods aim to decompose the set of objects into a set of disjointed clusters where the user predefines the resulting number of clusters. The k -means algorithm and the k -medoids algorithm are the most classical and the most commonly used partition-based clustering methods. Compared with the k -means, the k -medoids algorithm eliminates sensitivity to outliers, applicable only to small datasets because of its high computational complexity. Partitioning Around Medoids (PAM) algorithm realizes k -medoids clustering iteratively and greedily, i.e., in the iterative process, the greedy strategy is adopted to improve clustering quality by setting the maximum number of iterations. PAM

works efficiently for small datasets but does not scale well for large datasets [29].

To deal with more massive datasets, Kaufman and Rousseeuw (2008) proposed a sampling-based PAM algorithm—CLARA (Clustering LARge Applications)—which solved the PAM algorithm's problem in big data processing [30]. Instead of considering the whole dataset, CLARA uses a random sample and then applies the PAM algorithm to compute the best medoids from the sample. After repeated sampling, CLARA builds clusterings from multiple random samples and returns the best clustering as output. Algorithm 1 displays the ETC customer clustering procedure using the CLARA algorithm.

The distance from every non-medoids object O_j to different medoids M_f ($f = 1, 2, \dots, k$), represented as $d(O_j, M_f)$ is measured by Euclidean distance in the CLARA algorithm, as shown in the following:

$$d(O_j, M_f) = \sqrt{\sum_{u=1}^3 (o_{ju} - m_{fu})^2} \quad (3)$$

where u represents the index dimension of ETC customer and o_{ju} and m_{fu} denote the corresponding dimension values of O_j and M_f .

Input:

D - ETC customer index dataset and their associated class labels;
 $minbucket$ - the minimum number of observations in any terminal (leaf) node.

Output:

A decision tree of ETC customer segmentation.

Method:

- (1) create a node N ;
- (2) set a split point, a , for a specific segmentation index A , and split D into subsets D_1 and D_2 . Thus, for ETC segmentation index, three set of subsets are obtained;
- (3) computerize the Gini indexes of three indexes in dataset D , respectively. Determine an optimal splitting index;
- (4) repeat steps (1)–(3) until the samples in the subset are too few or the reduction of “node impurity” cannot be below the given threshold and create a leaf node;
- (5) the leaf node is labelled with the majority class in D to node N , and generate a decision tree of ETC customer segmentation;
- (6) select different subtrees (branches) in the decision tree and prune it by the cross-validated error and cost complexity;
- (7) output an optimal decision tree of ETC customer segmentation.

ALGORITHM 2: CART algorithm.

The actual distance $d(O_j, M)$ from the sample O_j to its cluster medoid is the minimum value in k distances:

$$d(O_j, M) = \min \{d(O_j, M_f), f \in (1, 2, \dots, k)\} \quad (4)$$

To determine whether current k -medoids are optimal, the average dissimilarity of this clustering, i.e., the arithmetic mean of distances from all samples in the dataset to their cluster medoid needs to be calculated, as shown in the following equation:

$$D_{average} = \frac{1}{N} \sum_{j=1}^N d(O_j, M) \quad (5)$$

where $D_{average}$ is the average dissimilarity and N is the number of samples in the ETC customer index dataset.

2.3. Segmentation Rules. After clustering analysis, each ETC customer is assigned a specific class label. Decision tree induction is the learning of decision trees from the class-labelled training dataset. The decision tree can be converted to classification of “IF-THEN” rules by tracing the path from the root node to each leaf node in the tree.

The most widely used decision tree algorithms are ID3 (Iterative Dichotomiser 3), C4.5 (a successor of ID3), and CART (Classification And Regression Trees). Compared to other decision tree algorithms, the CART algorithm simplifies the information theory based entropy model, while still retaining the entropy model’s advantages using a binary tree instead of a multi-way tree and the Gini index instead of the information gain ratio [31]. This study uses the CART algorithm to induce ETC customers’ segmentation decision tree, and Algorithm 2 shows the detailed procedure.

In the process of splitting, the Gini index measures the impurity of D or a data partition, as

$$Gini(D) = \sum_{i=1}^k p_i (1 - p_i) = 1 - \sum_{i=1}^k p_i^2 \quad (6)$$

where p_i is the probability that the sample in D belongs to class C_i ; k is the number of the class label in D .

If a binary split in segmentation index A partitions D into D_1 and D_2 , the Gini index of D , given that partitioning is as follows:

$$Gini_A(D) = \frac{|D_1|}{D} Gini(D_1) + \frac{|D_2|}{D} Gini(D_2) \quad (7)$$

The reduction in impurity that would be incurred by a binary split on segmentation index A is the following:

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (8)$$

The index that maximizes reduction in impurity (or, equivalently, has the minimum $Gini_A(D)$) is selected as the splitting attribute.

To extract rules from a decision tree, one rule is created for each path from the root to a leaf node. Each splitting criterion along a given path is logically “ANDed” to form the rule antecedent (“IF” part). The leaf node holds the class prediction, forming the rule consequent (“THEN” part).

2.4. Modeling Procedure. The modeling procedure for ETC custom segmentation includes the following steps.

(1) *Data Preprocessing and Index Extraction.* This step includes the following: cleaning raw ETC data and extracting customer segmentation indexes; selecting data subset and forming the

TABLE 3: Extraction results of ETC customer segmentation index.

ETC card No.	R	F	M
61010922230010* * * * *	30	1486	22055
61021101230035* * * * *	1399	44	1535
61021101230035* * * * *	1087	9	100
61032607230050* * * * *	56	119	4400
.....			

Note. To protect privacy, the ETC card's last six digits were replaced with asterisks (*).

ETC customer index dataset by setting a threshold value for each index.

(2) *ETC Customer Clustering*. This step includes the following: performing clustering analysis for the ETC customer index dataset and obtaining clustering results of the ETC customer.

(3) *Segmentation Rules Extraction*. This step includes the following: learning the decision tree of segmentation rules from the ETC customer index dataset (training tuple) and clustering results (class label) with the CART algorithm; extracting rules from the tree and realizing the final star-rating of the ETC customer.

Figure 1 displays the complete modeling procedure for ETC customer segmentation.

3. Results

3.1. Data Preprocessing and Index Extraction. In this study, the 2014 annual ETC data, over 31 million, of passenger vehicles with seven seats or fewer in Shaanxi province was chosen as basic data. First, the data were cleaned. Irrelevant data (toll-free vehicles) or abnormal passing data (for instance, entrance time is later than exit time) were deleted. Then 324,585 groups of ETC customer segmentation index data were extracted with the specified time $T_{set} = \text{"2015-1-2 00:00:00."}$ Table 3 shows the specific format.

Next, Figures 2(a)–2(c) show probability density distributions of three kinds of segmentation indexes. Further analysis indicates that when $R \leq 2160$, the percentage of ETC customers, who had consumption records within 90 days (2160 h) from the specified time, accounts for about 85% of the total. In the case of $F < 6$, that is, ETC customers with annual travel frequency of less than six times account for about 13.3%. In the case of $M < 200$, that is, ETC customers with annual monetary payments of less than 200 yuan account for about 18.6%, and those with more than 12,000 yuan account for about 0.77%.

To optimize the selected data subset and improve clustering accuracy, we filtered ETC customers who had too low frequency or extreme monetary values during data preprocessing. The filter criterion was $(F < 6) \cup (M < 200) \cup (M > 12000)$. Finally, an ETC customer index dataset containing 255,316 groups of ETC customers was formed.

Because of the massive data volume, a 2% random sampling was used to draw the scatter plot of "Frequency-Monetary", shown in Figure 3. The slope of the oblique line is 5, representing the single average toll of 5 yuan. Due to actual

FIGURE 1: Modeling procedure for ETC customer segmentation.

tolls being integral multiples of 5 yuan, the single average toll should be more than or equal to 5 yuan for normal payment vehicles (the slope is greater than or equal to 5). Figure 3 demonstrates that abnormal data generated by toll-free vehicles has been cleaned.

3.2. Clustering Results. In this study, the optimal number of clusters in ETC customer index dataset was estimated by the optimum average silhouette width [32]. The average silhouette method computes the average silhouette width of all customer samples for different values of k , and the optimal number of clusters k is the one that maximizes the average silhouette width over a range of possible values for

FIGURE 2: Distribution of ETC customer segmentation index.

FIGURE 3: Scatter plot of “Frequency-Monetary” (sampling data).

k . The calculation indicates that $k = 3$ corresponded to the maximum width, so the optimal number of clusters is 3.

By considering filtered ETC customers during data preprocessing, the above-mentioned three types of ETC customers were expressed by C_2 , C_3 , and C_4 . Filtered ETC customers $(F < 6) \cup (M < 200)$ and $(M > 12000)$ were, respectively, expressed as C_1 and C_5 .

Due to its vast amount of data, methods like k -means, PAM, and so forth are unable to realize the whole-sample clustering of the ETC customer index dataset. In CLARA algorithm, the bigger the number of samples (*samples*) and observations (*sampsize*) is preset, the more accurate clustering results will get, but the corresponding computational expense will also increase.

By presetting different combined parameters (*samples* and *sampsize*) and executing iterative computation on ETC customer index dataset with the CLARA algorithm, the comparison results of the optimal clustering medoids and run-time (s) were obtained, as listed in Table 4.

Table 4 indicates that the clustering medoids tend to converge with the increase of *samples* and *sampsize*. Taking data volume and time effectiveness into consideration, 2% ETC customers (*sampsize* = 5000) were selected randomly at each sampling, then iteratively running ten times (*samples* = 10) with CLARA algorithm and finally getting the class label of each ETC customer.

3.3. Segmentation Results. Presetting the minimum number of observations in any leaf node at $\text{minbucket} = 1000$, a “segmentation index–customer classification (C_2 , C_3 , and C_4)” decision tree was built using the CART algorithm, as shown in Figure 4.

TABLE 4: Calculation results of clustering medoids under different combined parameters.

cluster	<i>samples</i> = 5 <i>sample</i> = 500			<i>samples</i> = 10 <i>sample</i> = 5000			<i>samples</i> = 20 <i>sample</i> = 10000		
	<i>R</i>	<i>F</i>	<i>M</i>	<i>R</i>	<i>F</i>	<i>M</i>	<i>R</i>	<i>F</i>	<i>M</i>
C_2	3976	37	795	3750	27	765	3630	28	780
C_3	274	44	1025	316	43	995	302	43	1005
C_4	73	179	4340	176	160	3799	160	160	3877
run-time	0.55 s			59.69 s			846 s

FIGURE 4: Decision tree of ETC customer segmentation.

This decision tree contains six leaf nodes. The first line in each node displays the final fitted classification of observations (ETC customers), the second line shows the probability per classification (C_2 , C_3 , and C_4), the third line displays the total percentage of observations (ETC customers) in this node, and the sum across all leaves is 1.

Segmentation rules of ETC customers C_2 , C_3 , and C_4 were extracted from Figure 4, and the filtering rules of C_1 and C_5 were also incorporated. Finally, all were transformed into a set of “IF-THEN” segmentation rules, as listed in Table 5.

All ETC customers were classified as C_1 – C_5 , corresponding to different stars. Different star-rating customers and their summarized details are listed in Table 6.

Table 6 indicates that 324,585 ETC customers in Shaanxi province annually paid tolls 23.13 million times, with a total consumption of 546 million yuan in 2014. According to the current 5% favorable discount rate, the actual annual ETC toll revenue was 519 million yuan.

One-star customers accounted for only about 20.57%, with a total consumption contribution of 1.33%. For such customers, a strengthened propaganda and guidance plan should be drawn to improve their ETC usage rates. Two-star customers accounted for 8.15%, with a total consumption contribution of 4.71%. According to their important characteristics, such customers should be cultivated to tap

ETC payment potential. Three- and four-star customers accounted for 49.42% and 21.09%, respectively. The sum of consumption contributions was over 85%, indicating major customers in ETC service. In the future, an additional discount rate might be considered to build and enhance their self-worth. Five-star customers accounted for only 0.77%, but they contributed 7.6% to the total consumption, illustrating that they are key ETC customers. Thus, customizing a larger, additional discount rate for them is advisable. Meanwhile, their personal feelings about using the ETC system should be tracked and responded to, in order to achieve maximum improvement in the ETC service quality.

4. Discussion

An ETC customer segmentation index system was defined based on the RFM model. According to future operational requirements, new segmentation indexes can be introduced and adjusted with different weights, making star-rating results more suitable to the “Customer Pyramid” model [33].

There are differences in toll standards and usage characteristics among vehicle types. In this study, only ETC customers driving passenger vehicles with seven seats or fewer were segmented. Segmentation studies on other vehicle

TABLE 5: Segmentation rules of ETC customers.

Rule No.	Rule antecedent		Rule consequent	
R1	IF	$(F < 6) \text{ OR } (M < 200)$	THEN	Customer = C_1
R2	IF	$M > 12000$	THEN	Customer = C_5
R3	IF	$(200 \leq M < 2516) \text{ AND } (R \geq 1946)$	THEN	Customer = C_2
R4	IF	$(6 \leq F < 76) \text{ AND } (2516 \leq M < 3328)$	THEN	Customer = C_3
R5	IF	$(6 \leq F < 76) \text{ AND } (3328 \leq M \leq 12000)$	THEN	Customer = C_4
R6	IF	$(2516 \leq M \leq 12000) \text{ AND } (F \geq 76)$	THEN	Customer = C_4
R7	IF	$(200 \leq M < 2516) \text{ AND } (R < 1946) \text{ AND } (6 \leq F < 130)$	THEN	Customer = C_3
R8	IF	$(200 \leq M < 2516) \text{ AND } (R < 1946) \text{ AND } (F \geq 130)$	THEN	Customer = C_4

TABLE 6: Star-rating results for ETC customers.

Star	Number of customers	Total frequency	Total consumption (thousand yuan)	Ratio of customers (%)	Ratio of total consumption (%)
One-star	66765	384276	7240.2	20.57%	1.33%
Two-star	26457	963023	25727.6	8.15%	4.71%
Three-star	160405	7394715	170657	49.42%	31.25%
Four-star	68454	13187769	300927.5	21.09%	55.11%
Five-star	2504	1204661	41514.4	0.77%	7.60%

types should be conducted by following the proposed method combined with specific travel characteristics.

5. Conclusions

Applying big data technology, this study proposed an ETC customer segmentation method. Segmentation indexes were extracted from ETC data, customer clustering analysis was performed based on the CLARA algorithm, and segmentation rules were created. In this case study, ETC customer segmentation and star-rating were realized and travel characteristics and service strategies for each customer type were analyzed. The study thus provides an innovative idea for implementing precision marketing and creating hierarchical discount rates for ETC customers. Meanwhile, the study also provides theoretical support for further increase in the ETC customer scale and payment ratio, for an improved level of decision-making in expressway operation and management.

Data Availability

Data is not authorized; the authors regret that it is not available.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Natural Science Basic Research Plan in Shaanxi Province of China (Grant no. 2016JM5052) and the Fundamental Research Funds for the

Central Universities of Ministry of Education of China (Grants nos. 310821173102 and 300102218203).

References

- [1] Ministry of Transport of China, in *Proceedings of the third regular news conference in 2017*, Beijing, China, 2017, <http://www.mot.gov.cn/2017wangshangzhibo/2017third/>.
- [2] K. A. Richards and E. Jones, "Customer relationship management: Finding value drivers," *Industrial Marketing Management*, vol. 37, no. 2, pp. 120–130, 2008.
- [3] Z. Soltani and N. J. Navimipour, "Customer relationship management mechanisms: A systematic review of the state of the art literature and recommendations for future research," *Computers in Human Behavior*, vol. 61, pp. 667–688, 2016.
- [4] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3964–3973, 2012.
- [5] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert Systems with Applications*, vol. 31, no. 1, pp. 101–107, 2006.
- [6] H. Hwang, T. Jung, and E. Suh, "An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry," *Expert Systems with Applications*, vol. 26, no. 2, pp. 181–188, 2004.
- [7] I. Benitez, A. Quijano, J. L. Diez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 437–448, 2014.
- [8] J. J. López, J. A. Aguado, F. Martín, F. Muñoz, A. Rodríguez, and J. E. Ruiz, "Hopfield-K-Means clustering algorithm: A proposal for the segmentation of electricity customers," *Electric Power Systems Research*, vol. 81, no. 2, pp. 716–724, 2011.

- [9] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," *Electronic Commerce Research and Applications*, vol. 10, no. 3, pp. 331–341, 2011.
- [10] C.-F. Tsai, Y.-H. Hu, and Y.-H. Lu, "Customer segmentation issues and strategies for an automobile dealership with two clustering techniques," *Expert Systems with Applications*, vol. 32, no. 1, pp. 65–76, 2015.
- [11] L. F. Wei, "Design and Implementation of Airline Customer Segmentation System Based on Ant Colony Clustering Algorithm," in *Proceedings of the International Conference on Materials Science and Information Technology (MSIT 2011)*, Singapore, 2011.
- [12] T. Teichert, E. Shehu, and I. von Wartburg, "Customer segmentation revisited: The case of the airline industry," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 227–242, 2008.
- [13] W.-Y. Chiang, "Discovering customer value for marketing systems: an empirical case study," *International Journal of Production Research*, vol. 55, no. 17, pp. 5157–5167, 2017.
- [14] Y.-H. Cheng and T.-Y. Huang, "High speed rail passenger segmentation and ticketing channel preference," *Transportation Research Part A: Policy and Practice*, vol. 66, no. 1, pp. 127–143, 2014.
- [15] B. Zhang and Q.-Y. Peng, "Railway Freight Customer Segmentation Based on KFAV Model," *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology*, vol. 17, no. 3, pp. 235–242, 2017.
- [16] Y. Zhong and Y. Guo, "Research of applying data mining in segmentation of railway freight customers," *Beijing Jiaotong Daxue Xuebao/Journal of Beijing Jiaotong University*, vol. 32, no. 3, pp. 25–36, 2008.
- [17] L. Duan, J. Rezaei, L. Tavasszy, and C. Chorus, "Heterogeneous valuation of quality dimensions of railway freight service by Chinese shippers choice-based conjoint analysis," *Transportation Research Record*, vol. 2546, pp. 9–16, 2016.
- [18] W.-H. Lee, S.-S. Tseng, and C.-H. Wang, "Design and implementation of electronic toll collection system based on vehicle positioning system techniques," *Computer Communications*, vol. 31, no. 12, pp. 2925–2933, 2008.
- [19] W.-Y. Shieh, C.-C. Hsu, S.-L. Tung, P.-W. Lu, T.-H. Wang, and S.-L. Chang, "Design of infrared electronic-toll-collection systems with extended communication areas and performance of data transmission," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 25–35, 2011.
- [20] D. Levinson and E. Chang, "A model for optimizing electronic toll collection systems," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 4, pp. 293–314, 2003.
- [21] H. Yamazaki, N. Uno, and F. Kurauchi, "The effect of a new intercity expressway based on travel time reliability using electronic toll collection data," *IET Intelligent Transport Systems*, vol. 6, no. 3, pp. 306–317, 2012.
- [22] V. Astarita, M. Florian, and G. Musolino, "A microscopic traffic simulation model for the evaluation of toll station systems," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 692–697, Oakland, CA, USA, August 2001.
- [23] M. L. Zarrillo and A. E. Radwan, "Methodology SHAKER and the capacity analysis of five toll plazas," *Journal of Transportation Engineering*, vol. 135, no. 3, pp. 83–93, 2009.
- [24] K. Ozbay and M. Yildirimoglu, "Comparison of real-time travel time estimation using two distinct approaches: Universal kriging and mathematical programming," in *Proceedings of the 14th IEEE International Intelligent Transportation Systems Conference, ITSC 2011*, pp. 1083–1088, Washington, DC, USA, October 2011.
- [25] H. Yang, K. Ozbay, and K. Xie, "Improved travel time estimation for reliable performance measure development for closed highways," *Transportation Research Record*, vol. 2526, pp. 29–38, 2015.
- [26] A. Nairn and P. Berthon, "Creating the Customer: The Influence of Advertising on Consumer Market Segments - Evidence and Ethics," *Journal of Business Ethics*, vol. 42, no. 1, pp. 83–99, 2003.
- [27] A. Hughes, *Strategic Database Marketing*, McGraw-Hill Education, New York, NY, USA, 1994.
- [28] B. Stone and R. Jacobs, *Successful Direct Marketing Methods*, McGraw-Hill Education, New York, NY, USA, 8th edition, 2007.
- [29] J. W. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Waltham, MA, USA, 3rd edition, 2011.
- [30] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, Hoboken, NJ, USA, 2008.
- [31] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, Chapman and Hall/CRC, Boca Raton, FL, USA, 1984.
- [32] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [33] V. A. Zeithaml, R. T. Rust, and K. N. Lemon, "The customer pyramid: Creating and serving profitable customers," *California Management Review*, no. 4, pp. 118–142, 2001.