Proceedings of Engineering and Technology Innovation, vol. 1, 2016, pp. 15 - 18

The Application of Target Analysis in Electricity Demand-Side Management

Fung-Fei Chen[1], Seng-Cho Chou[2, *], Chiao-Yi Wang[2], Tai-Ken Lu[3]

[1]Taiwan Power Company Research Institute, Taipei, 100, Taiwan

[2]Department of Information Management, National Taiwan University, Taipei, 100, Taiwan

[3]Department of Electrical Engineering, National Taiwan Ocean University, Keelung, 202, Taiwan

Received 13 April 2015; received in revised form 11 May 2015; accepted 18 June 2015

Abstract

Recently, target analysis combined with database technology and data mining has been widely used in industries such as marketing, finance, insurance, telecommunications, advertising, and e-commerce. Because of the unique complexities of user behavior in electricity demand, examples of target analysis applications have yet to be seen. Considering the industry's urgent need to enhance the efficiency of electricity demand-side management, this study aims to build a mining analysis model for potential target users of interruptible load that both fully reflects consumer behavior characteristics and serves as a rule for static comparisons. The results of a data mining analysis of the Taiwan Power Company (Taipower)'s interruptible loads 1 to 6 show that the number of potential target users is 1669, which is 21% of the original mining population. Additionally, the target users who were classified to have "the most potential" for all categories of interruptible load only accounted for 0.76% of the total mining population (= 59/7814), verifying the mining effects.

Keywords: Target analysis, data mining, association rule, electricity demand-side management

1. Introduction

In management, target analysis has always been crucial to enterprise operations because it enables full marketing exploitation of the 80/20 rule and increases the opportunities for product recommendation and cross selling. Another aspect of target analysis that has attracted significant attention and extensive discussions in recent years is its ability to identify valuable potential target users. Target analysis coupled with database technology and data mining, which emerged from analysis of significant amounts of data, is currently applied extensively in the fields of marketing, finance, insurance, manufacturing, and medical care [2, 4, 13, 16]. By employing target analysis data mining, marketers can obtain hidden knowledge or identify the most commercially valuable information from significant amounts of unsorted data. When used appropriately, target analysis can provide significant overall benefits for the organization.

Most previous studies employed decision tree analysis or data mining through artificial neural networks to identify and extract the characteristics of potential target users, before establishing comparison rules to enable companies to examine potential customers as marketing targets [3, 5, 6, 8, 10, 12, 19]. The purpose of these investigations was to determine the consumer behavior characteristics of existing customers, and then use these characteristics to identify hidden potential target customers. However, for industries where consumer behavior is influenced by industry policies or other external factors, because existing customers do not exhibit the same consumer behavior as potential customers, establishing a comparison rule between the existing customers and potential customers is difficult, resulting in a lack of research in this area. This phenomenon is particularly common in telecommunication and utility industries, where marketing measures may adversely affect the consumer behaviors of existing users, decreasing the objectivity of the consumer behavior characteristics identified using target analysis, further hindering comparisons [1, 14, 18].

To address these issues and compensate for the disadvantages of current target marketing analysis methods, we conduct a case study of Taipower's demand-side management data mining analysis of potential target users. This study also constructs a mining analysis model for potential target users, which both fully reflects consumer behavior characteristics and serves as a rule for static comparisons. We hope that the target analysis mining model proposed in this study can provide a list of the most valuable potential users of an interruptible load. The analysis model is also expected to provide a useful reference for identifying potential target users for other purposes related to electricity demand, thereby facilitating the demand-side management of electricity.

2. Experiment

2.1. Research design

Taipower's demand-side management measures include time use rates, seasonal rates, interruptible loads, and air conditioning control; among which, interruptible load is the most effective for reducing peak power loads. Since the introduction of Taipower's interruptible load policy in 1987, despite the seven interruptible load categories, by the end of 2005, the number of interruptible load users was only 670. Compared to the target population of over 10,000 qualified users, the number of actual users is minimal, presenting significant scope for promotion. Therefore, target analysis and mining analysis of significant amounts of user data must be employed to identify potential target users and expand the scope of promotion [9].

Target analysis generally includes customer data collection, data analysis, the extraction of segmentation variables, and the establishment of segmentation market characteristics [7]. Without identical consumer behavior data and uncertainty whether basic variables are sufficient for market segmentation, the consumer behavior of existing customers can be used as the foundation for segmentation to identify the basic variable characteristics of each market segment. However, applying the existing target marketing analysis method, which selects the differing basic variables of the segmentation market for characteristic analysis, may lead to losses of consumer behavior-related data or deficiencies in the descriptive variable dimensions, thereby preventing the effective convergence

*Corresponding author. Email address: chou@ntu.edu.tw

of the target marketing scope. To solve this problem, we determine the behavioral data of existing customers who have changed because of promotional measures, identify the basic variables that are not subject to influence but are related, and eventually establish potential target customer rational comparison rules as the target analysis method in this study. The target marketing analysis model based on a lack of identical consumer behavior data is shown in Fig. 1.

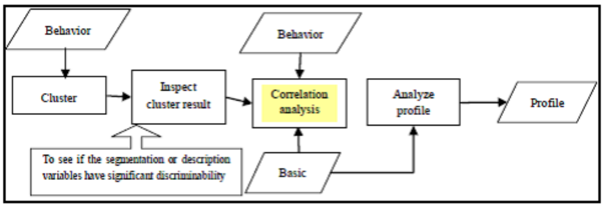

Fig. 1 The target marketing model used when there is insufficient information about equivalent consumption behavior

2.2. Data description

This study primarily examines the data of low-voltage users who qualify for Taipower's interruptible load policies, categorizing them as either usage data or basic data. Designed to record the various electricity consumption behaviors of existing users who adopt interruptible load policies, data usage includes 79 variables written in numerical form, excluding "electricity no." and "industry." To ensure the quality of this research, we first omit a number of the insignificant columns based on the experience of experts. Then, to facilitate subsequent data analysis, we categorize the remaining 51 fields into five electricity consumption factors according to their features and relevance.

(1) "Frequent and Peak" demand or reading group: All 16 variables show users' frequent electricity consumption behavior during peak hours.

(2) "Off-Peak" demand or reading group: All 8 variables show users' electricity consumption behavior during off-peak hours.

(3) "Saturday Semi-peak" demand or reading group: All 8 variables show users' electricity consumption behaviors during semi-peak hours.

(4) "Comparative" demand or reading group: All 11 variables show users' varying electricity consumption behaviors in different periods.

(5) "Others" group: The remaining 8 variables.

Additionally, of the 38 basic data variables, including both classifying and analysis variables, a number of the variables that were irrelevant to electricity consumption behavior were omitted based on expert experience. Data of the 7 remaining variables were used to develop comparison rules for identifying the potential users who may accept interruptible load policies, as shown below.

Table 1 The relevant variables of users' basic data

Attributes	Column Name	Column Description
Non-numeric	TR_CODE	by industry
	CO_CODE	by contract
	EL_CODE	by usage
	CS_THREE_TOU	three kinds of interval contract customer
Numeric	CS_CAPACITY	contract capacity
	CS_UP_CAPACITY	off-peak contract capacity
	CS_NS_CAPACITY	half-peak contract capacity

By examining the electricity consumption behaviors of existing users with various interruptible load policies and segmenting their objective effectiveness or determining the amount they can generate, we aim to identify the customers with the most potential, the second most potential, and the least potential among the users of various interruptible loads. This study also correlates the basic files of existing users and extracts the unique characteristics from the basic data of this beneficial user type. A comparison rule for identifying potential target users is then developed to narrow the target marketing scope and best employ the limited time and cost resources.

2.3. Data mining process

Because of the requirements of the analysis model and the characteristics of relevant data adopted in this study, we employ the following techniques for data mining analysis: principal component analysis, data clustering, correlation analysis, association rule and MANOVA tests. Thus, a data mining analysis process is established, as shown in Fig. 2.

2.4. Results

The results of all data mining stages are presented below.

2.4.1. Analysis of data dimension reduction

To determine the most suitable linear equations, this study adopted principal component analysis. By condensing the variables into a representative indicator, the research complexity and dimensions were reduced effectively, facilitating the smooth operation of subsequent research processes. The principal component analysis equation is as follows:

$$\begin{aligned} Y_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ &\vdots \\ Y_p &= \mathbf{a}_p'\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p \end{aligned} \tag{1}$$

where Y_1 to Y_p are representative principal component indicators extracted from a certain factor dimension. Depending on the difference of internal data, each factor dimension may possess 1 top (the number of variables in that dimension) representative indicators. A suitable critical threshold (cumulative explanatory variance) can be determined for the selection.

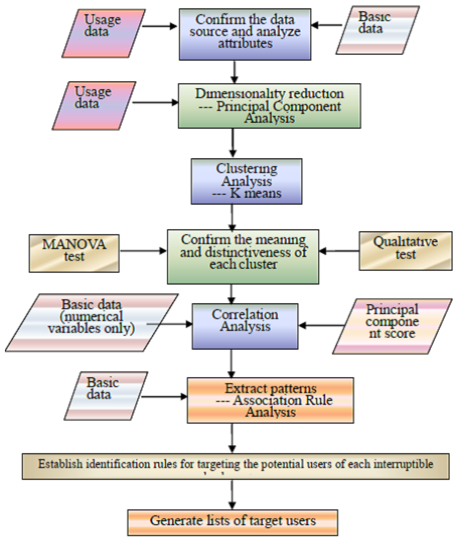

Fig. 2 Data analysis process

After repeated verification, we found that principal component analysis of dimensions 4 and 5 does not provide good results and fails to converge dimensions effectively. To ensure the credibility, reliability, and explanatory power of the data analysis results, only the first representative principal component indicator (PC1 is the first representative principal component indicator of dimension 1, PC2 is the first representative principal component indicator of dimension 2, etc) of the first three usage data dimensions mentioned in Section 2.4 were used as segmentation variables for future data collection, as shown in Table 2.

Table 2 Cumulative explanatory variance of representative principal component indicators

Representative Principal Component Indicator	Interruptible Load1	Interruptible Load2	Interruptible Load3	Interruptible Load4	Interruptible Load5	Interruptible Load6
	Cumulative Explanatory Variance					
PC1	64.37%	57.01%	78.26%	76.98%	49.69%	77.19%
PC2	98.19%	96.14%	99.02%	93.72%	95.62%	99.70%
PC3	95.48%	95.35%	98.51%	93.06%	92.06%	99.64%

2.4.2. Cluster analysis

The most commonly used data clustering methods include hierarchical clustering, non-hierarchical clustering, and artificial neural networks[15]. Among them, the K means method provides superior clustering results and higher efficiency compared to hierarchical clustering and artificial neural network clustering [11, 15, 17]. Using a two-dimensional and three-dimensional scatter plot and chi-square plot, we verified that the usage data in this study has normal distribution. After assessments and experimentation, the K means method was employed for cluster analysis. Tables 3 and 4 show the cluster analysis results of the data of existing users.

2.4.3. MANOVA Test

A MANOVA test was conducted in this study to determine whether the clustering results can differentiate. Using Wilk's lambda, all impact factors (i.e., segmentation variables) were examined to

Table 3 Cluster analysis results-the number of existing users in each cluster

Cluster	Number of Existing Users in Each Cluster					
	Interruptible Load 1	Interruptible Load 2	Interruptible Load 3	Interruptible Load 4	Interruptible Load 5	Interruptible Load 6
1	6	2	1	5	17	1
2	23	5	9	5	4	1
3	101	31	134	12	44	34

Table 4 Cluster analysis results – the number of interruptible load users belonging to more than one cluster

Interruptible Load measure	Number of participants who belong to more than one cluster	Number of participants in each measure	Ratio of participants who belong to more than one cluster in each measure
1	4	126	0.0317
2	1	37	0.0270
3	1	143	0.0070
4	1	21	0.0476
5	2	63	0.0317
6	0	36	0.0000

Determine whether they had a significant influence on the overall segmentation results. If the likelihood of the Wilk's lambda value being smaller than a certain critical chi-square distribution value is minimal (<0.0001), we can conclude that the impact factor has substantial influence (i.e., discriminatory ability), as shown in Eq. (2). From a quantitative analysis perspective, if each impact factor has significant influence, the credibility of the cluster analysis results can be verified, as shown in Table 5.

Table 5 MANOVA test results for interruptible load 1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilk's Lambda	0.1200576	390.41	6	1242	<0.0001
Pillai's Trace	0.9076786	172.29	6	1244	<0.0001
Hotelling-Lawley Trace	7.0983087	734.09	6	826.22	<0.0001
Roy's Greatest Root	7.0656117	1464.94	3	622	<0.0001

$$
\begin{aligned}
& H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \, (no \quad factor 1 \quad effects) \\
& H_1 : At \quad least \quad \text{one} \quad \tau_i \neq 0 \\
& Let \quad \Lambda^* = \frac{|SSP_{res}|}{|SSP_{fac1} + SSP_{res}|} \\
& For \quad large \quad samples, \quad reject \quad H_0 \quad if \\
& -\left[gb(n-1) - \frac{p+1-(g-1)}{2} \right] \ln \Lambda^* \succ \chi^2_{(g-1)p}(\alpha)
\end{aligned} \tag{2}
$$

2.4.4. Correlation analysis

Because of the changing use behavior of existing interruptible load users, basic data were used to analyze user characteristics. To ensure that the basic variables after correlation reflect both the electricity consumption behavior of all users and the results of cluster analysis, correlation analysis of the three analytical variables in Table 2 should be performed and the segmentation variables employed for cluster analysis (Table 6). Suitable variables identified in the analysis can be used for further mining analysis. An example of the test results for interruptible load 1 is shown below.

Table 6 Correlation analysis results for interruptible load 1

	PC1	PC2	PC3
CS_CAPACITY	-0.9130	-0.9675	-0.9631
CS_UP_CAPACITY	-0.3037	-0.5418	-0.5148
CS_NS_CAPACITY	0.0436	0.0433	0.0417

Of the analytical variables of interruptible load 1, CS_CAPACITY and CS_UP_CAPACITY are the most relevant to users' consumption behaviors; thus, they can be included as the descriptive variables of interruptible load 1.

2.4.5. Association rule analysis

The association rule was employed to establish data correlation rules. After computations to determine whether it satisfies the threshold limit of minimal support and minimal confidence, the rule is matched with the Apriori algorithm to select appropriate correlation rules.

Data Analysis Results: An example of the representative characteristic behaviors and correlation mining threshold of users with "the most potential" for interruptible load 1 is shown below.

- Number of Existing Users: 6
- Minimal Support; 6
- Minimal Confidence: 0.8
- Importance: 0.2
- Essential Association Rule:

For users with an "A" contract type, the likelihood that they are from the top 20 industry types = 11 (basic metal industries) is > 0.8.

For three-stage users in Category 2, the likelihood that they are from the top 20 industry types = 11 (basic metal industries) is > 0.8.

For users whose usage type = 5, the likelihood that they are three-stage users in Category 2 is > 0.8.

For users whose usage type = 5, the likelihood that their contract type = 5 is > 0.8.

When the necessary correlation rules for all clusters are condensed into rigorous comparison rules, the number of potential target users and existing users of every cluster can be identified, as shown in Table 7.

The total number of potential target users shown in Table 7 is 4,687. However, because a number of them overlap between clusters, the actual number of potential target users is 3,770. Additionally, some of the clusters have few existing users (no more than 10, and some as few as 1 or 2); thus, the comparison rules established with these clusters are relatively loose and can result in an overestimation of potential target users. By contrast, potential target users mined from clusters with more existing users yield superior convergence. The potential target users from these clusters (excluding clusters with only 1 or 2 existing users) totals 2,058; excluding overlapped users, the actual number of potential target users is 1,669, which is 21% of the original mining population. The number of target users with the most potential from all interruptible load policies is 0.76% of the overall mining population (= 59/7814).

Table 7 Correlation rule comparison analysis results for all clusters

Cluster	Interruptible Load 1		Interruptible Load 2		Interruptible Load 3	
	Number of Potential Target Users	Number of Existing Users	Number of Potential Target Users	Number of Existing Users	Number of Potential Target Users	Number of Existing Users
1(Most Potential)	23	6	24	2	120	1
2(Second Most Potential)	0	23	22	5	4	9
3(Least Potential)	129	101	668	31	552	134

Cluster	Interruptible Load 4		Interruptible Load 5		Interruptible Load 6	
	Number of Potential Target Users	Number of Existing Users	Number of Potential Target Users	Number of Existing Users	Number of Potential Target Users	Number of Existing Users
1(Most Potential)	69	5	35	17	167	1
2(Second Most Potential)	33	5	8	4	2318	1
3(Least Potential)	64	12	101	44	350	34

2.4.6. Generating a list of potential target users

Using the correlation rules extracted from clusters, we compile a list of potential target users from among the customers not yet using interruptible load policies. Table 8 shows the first five results for users with the most potential in interruptible load 1.

Table 8 Interruptible load 1-users with the most potential

Electricity No.	Company Name	Frequent Contract Capacity(kW)
20157******	Gloria Material Technology Corp.	22992
04414******	Chungli Factory II. Minchali Metal Industry Co. Ltd.	18000
20662******	Tung Mung Development Co. Ltd. Hsueh-Chia Factory	13500
11599******	Yenwu Corp.	9600
06193******	Cheng Loong Corp.	4600

3. Conclusions

By applying target analysis to electricity demand-side management and mining analysis to a list of potential target users, this study develops a mining analysis model for potential target users of interruptible load that both fully reflects consumer behavior characteristics and serves as a rule for static comparisons. This model was designed to compensate for the deficiencies of the current target marketing analysis model. The results show that the mining analysis model proposed in this study can effectively narrow the target marketing scope to 21% of the overall mining population, with a condensing capability of 79%, which facilitates the segmentation and differentiation of customers with various potential benefits and value. The proportion of target users with the most potential for all interruptible load policies is 0.76% of the overall mining population (= 59/7814). Each step of the mining analysis process underwent thorough quantitative (theoretical verification) and qualitative testing (industry knowledge and experience-based assessments). The practical significance of this study is that it provides the electricity industry with information of the most potential and valuable target users. The mining analysis model proposed in this study can be employed by relevant marketing decision makers to identify potential target customers for other demand-side management policies and would have a lot to be referenced for utility in the future.

References

[1] A. Al-Ghandoor, et al. "Residential past and future energy consumption: potential savings and environmental impact," Renewable and Sustainable Energy Reviews, 2008.

[2] E. Bayam, J. Liebowitz, and W. Agresti, "Older drivers and accidents: a meta-analysis and data mining application on traffic accident data," Expert Systems with Applications, vol. 29, pp. 598-629, 2005.

[3] J. Z. Bloom, Market Segmentation. "A neural network application," Annals of Tourism Research, vol. 32, pp. 93-111, 2005.

[4] R. J. Brachman, T. Khabaza, W. Kloesgen, G. Piatetsky-Shapiro, and E. Simoudis, "Mining business databases," Communication of the ACM, vol. 39, pp. 42-48, 1996.

[5] S. W. Changchien, and T. C. Lu, "Mining association rules procedure to support on-line recommendation by customers and products fragmentation," Expert Systems with Applications, vol. 20, pp. 325-335, 2001.

[6] S. Daskalaki, I. Kopanas, M. Goudara, and N. Avouris, "Data mining for decision support on customer insolvency in telecommunications business," European Journal of Operational Research, vol. 145, pp. 239-255, 2002.

[7] S. Dibb, and P. Stern, "Questioning the reliability of market segmentation techniques," Omega International Journal of Management Science, vol. 23, pp. 625-636, 1995.

[8] S. H. Ha, and S. C. Park, "Application of data mining tools to hotel data mart on the Intranet for database marketing," Expert Systems with Applications, vol. 15, pp. 1-31, 1998.

[9] http://www.taipower.com.tw, 2012.

[10] S. Y. Hung, D. C. Yen, and H. Y. Wang, "Applying data mining to telecom churn management," Expert Systems with Applications, vol. 31, pp. 515-524, 2006.

[11] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," ACM Computing Surveys, vol. 31, pp. 264-323, 1999.

[12] T. S. Lee, C. C. Chiu, Y. C. Chou, and C. J. Lu, "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines," Computational Statistics and Data Analysis, vol. 50, pp. 1113-1130, 2006.

[13] D. R. Liu, and Y. Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," Information & Management, vol. 42, pp. 387-400, 2005.

[14] S. Roberts, "Demographics, energy and our homes," Energy Policy, vol. 36, pp. 4630-4632, 2008.

[15] S. S. Shahapurkar, and M. K. Sundareshan, "Comparison of self-organizing map with k-means hierarchical clustering for bioinformatics applications," Neural Networks, Proceedings. 2004 IEEE International Joing Conference on, pp. 1221-1226, 2004.

[16] W. E. Spangler, M. Gal-Or, and J. H. May, "Using data mining to profile TV viewers," Communication of the ACM, vol. 46, pp. 67-72, 2004.

[17] A. Ultsch, "Self-organizing neural networks perform different from statistical k-means clustering," Proc. Conf. Soc. For Information and Classification, Basel, 1995.

[18] W. O. Onuh, A. T. Valerio, A. Permalino Jr., "Residential demand for electricity dasmarinas, cavite, philippines," Journal of Global Business & Economics, vol. 2, pp. 1-22, 2011.

[19] C. H. Wu, S. C. Kao, Y. Y. Su, and C. C. Wu, "Targeting customers via discovery knowledge for the insurance industry," Expert Systems with Applications, vol. 29, pp. 291-299, 2005.

