Expert Systems with Applications 29 (2005) 145–152

Expert Systems with Applications

www.elsevier.com/locate/eswa

Intelligent profitable customers segmentation system based on business intelligence tools

Jang Hee Lee[a,*], Sang Chan Park[b,1]

[a]School of Industrial Management, Korea University of Technology and Education, 307 Gajeon-ri, Byeong cheon-myun, Cheonan City, Choongnam Province 330-708, South Korea

[b]Department of Industrial Engineering, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Kusong-dong, Yusong-gu, Taejon 305-701, South Korea

Abstract

For the success of CRM, it is important to target the most profitable customers of a company. Many CRM researches have been performed to calculate customer profitability and develop a comprehensive model of it. Most of them, however, had some limitations and accordingly the customer segmentation based on the customer profitability model is still underutilized. This paper aims at providing an easy, efficient and more practical alternative approach based on the customer satisfaction survey for the profitable customers segmentation. We present a multi-agent-based system, called the survey-based profitable customers segmentation system that executes the customer satisfaction survey and conducts the mining of customer satisfaction survey, socio-demographic and accounting database through the integrated uses of business intelligence tools such as DEA (Data Envelopment Analysis), Self-Organizing Map (SOM) neural network and C4.5 for the profitable customers segmentation. A case study on a Motor company's profitable customer segmentation is illustrated.

Keywords: Customer relationship management; Customer profitability; Customer segmentation; Customer satisfaction survey

1. Introduction

In today's competitive business environment, the ability to identify profitable customers, build their long-term loyalty and steadily expand existing relationships is key competitive factors to a company. To meet these factors, companies across a wide range of industries have made Customer Relationship Management (CRM) one of the leading business strategies, integrating sales, marketing and service across multiple business units and customer contact points.

CRM helps companies understand the value of customers, target their most profitable customers, cultivate and maintain high-quality relationships that increase loyalty and profits. Precise evaluation of customer profitability and targeting the most profitable customers are crucial elements for the success of CRM.

Many CRM researches have been performed to calculate customer profitability based on customer lifetime value and develop a comprehensive model of it. Most of them, however, had some limitations by not considering such as the change of profit contribution resulted from the customer defection (Berger & Nasr, 1998; Gupta & Lehmann, 2003). They need further extensions considering additional factors such as customer reactivation possibility, attracting/service cost and causes of customer defection.

On the other hand, the customer segmentation based on their profitability to a company is still an underutilized approach. This study aims at providing an easy, efficient and more practical alternative approach based on the customer satisfaction survey for the profitable customers segmentation instead of using a customer profitability model, which is an important tool for marketing and managing customer relationships by providing the information of overall satisfaction level, repurchase intentions, word-of-mouth intentions, etc.

In our approach, we use intelligent tools such as Data Envelopment Analysis (DEA), Self-Organizing Map (SOM)

* Corresponding author. Tel.: +82 41 560 1446; fax: +82 41 560 1439.
E-mail addresses: janghlee@kut.ac.kr (J.H. Lee), sangpark@kaist.ac.kr (S.C. Park).
[1] Tel.: +82 42 869 2960; fax: +82 42 869 5920.

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.01.013

neural network and C4.5 to segment profitable customers. DEA evaluates efficiency through the relation analysis between the company's input costs for a customer (e.g. marketing cost, production cost, inventory cost, delivery cost, service cost and relationship management cost) and the output (e.g. his/her satisfaction level, repurchase intentions and word-of-mouth intentions in the customer satisfaction survey and his/her profit contribution to it).

Through the successive mining of customer satisfaction survey and socio-demographic data by SOM and C4.5, we segment profitable customers among all the surveyed customers.

We present a survey-based profitable customers segmentation system (SPCSS) that designs, executes (on-line, e-mail, etc.) the customer satisfaction survey for all customers in customer database of a company and conducts those mining works for the profitable customers segmentation. SPCSS has an architecture based on intelligent agent technology and also the integration of those mining process into decision support system framework by means of applying that technology.

This paper is organized as follows. Section 2 presents a review of literature in the profitable customer segmentation and customer satisfaction survey. In Section 3, we introduce our research methodology for profitable customer segmentation based on customer satisfaction survey and the basic structure of proposed system incorporated that methodology is presented. A case study on a Motor company's profitable customer segmentation in South Korea is illustrated in Section 4 and the concluding remarks are presented in Section 5.

2. Profitable customer segmentation and customer satisfaction survey

Traditional customer segmentation models were based on demographic, attitudinal, and psychographic attributes of a customer (Griffin, 2003). They gave too simple results and poor accuracy for today's complicated business environment. Recently, the customer segmentation based on customer transactional and behavioral data (e.g. purchases type, volume and history, call center complaints, claims, web activity data, etc.) collected by various information systems is commonly used. However, the customer segmentation based on his/her profitability to a company is still underutilized.

Customer profitability is a customer-level measure that refers to the revenues less the costs which one particular customer generates over a given period of time and has been studied the name of Customer value, Customer Lifetime Value, LTV and Customer Equity. Many customer profitability researches focused on the future cash flow derived from the past profit contribution and did not considered the change of profit contribution resulted from the customer defection (Berger & Nasr, 1998; Gupta & Lehmann, 2003).

Hwang, Jung, and Suh (2004) suggested a new customer profitability model considering past profit contribution, potential benefit indicated cross-selling and up-selling opportunity, and defection probability of a customer measured customer loyalty and segmented customers based on their model. However, they said that it had some limitations such as not considering the reactivation possibility of customers, attracting/servicing cost and causes of customer defection.

It is difficult and complicated to develop an effective and exact customer profitability model and segment profitable customers based on that model. In this study, we provide an easy, efficient and more practical alternative approach through the customer satisfaction survey for the profitable customers segmentation instead of using that model.

The typical customer satisfaction survey collects data on the causal context of satisfaction, i.e. antecedents (e.g. perceived performance of various product attributes/service) and consequences (e.g. overall satisfaction level, repurchase intentions and word-of-mouth intentions). According to the Satisfaction-Profit Chain principle (Anderson & Mittal, 2000), improving product and service attributes causes increased customer satisfaction, increased customer satisfaction leads to greater customer retention and improving customer retention greater profitability.

Empirical Researches have shown that increasing overall satisfaction leads to greater repurchase intentions, as well as to actual repurchase behavior and companies with high customer satisfaction and retention can expect higher profits (Reichheld & Frederick, 1996).

In this study, we use the customer's overall satisfaction level, repurchase intentions, word-of-mouth intentions obtained from the customer satisfaction survey and his/her profit/loss to a company derived from the accounting database of it for the first step of profitable customers segmentation.

3. Profitable customers segmentation based on customer satisfaction survey

We propose a survey-based profitable customers segmentation system (SPCSS) based on data mining and agent technology that designs, executes (on-line, e-mail, etc.) customer satisfaction survey and conducts predefined mining processes for the profitable customers segmentation. SPCSS has a multi-agent based architecture and the integration of predefined mining processes into decision support system framework (Fig. 1).

There are three types of intelligent agents within the SPCSS architecture: Survey management (SM) agent with survey knowledge base that provides system co-ordination, facilitates (mined) knowledge communication, and takes the charge of design and execution of customer satisfaction survey, profitable customers segmentation (PCS) agent that segments profitable customers among all the surveyed

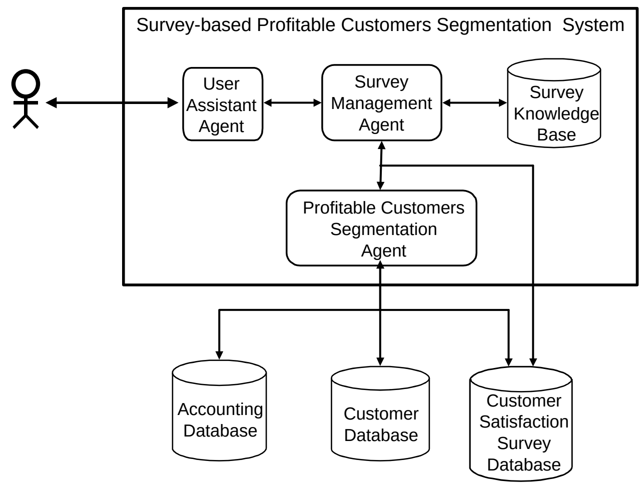

Fig. 1. Architecture of SPCSS.

customers through the mining of integrated data from the customer satisfaction survey and accounting database and decides the priority order for each non-profitable customer according to the size of possibility that he/she is converted to profitable one through the mining of integrated data from the customer satisfaction survey and customer database, and user assistant agent that acts as the intelligent interface agent between the user (e.g. the engineer of customer satisfaction center) and the SPCSS.

3.1. Profitable customers segmentation by the PCS agent

Fig. 2 shows the segmentation process of profitable customers among all the surveyed customers in PCS agent.

The first step is to find out the customers among all the surveyed ones that have higher efficiency about their output (e.g. the level of customer satisfaction, repurchase intentions, word-of-mouth intentions and profit/loss to the company) from company's input costs for them (e.g. marketing cost (campaign and advertisement), production cost, inventory cost, delivery cost, service cost and relationship management cost). We call the group of customers with higher efficiency HECG (High Efficiency Customer Group) in this study.

To find out HECG, PCS agent employs DEA (Data Envelopment Analysis), an efficiency measurement tool, to evaluate the cost efficiency of all the surveyed customers (Charnes, Cooper, Lewin, & Seiford, 1994). DEA evaluates their efficiencies through the relation analysis between the company's input costs for them and the output.

Because the customers who belong to HECG create more superior output than a company's input costs for them, they have an important effect on company's current and future profit generation. However, undesirable customers can belong to HECG because of the inaccuracy and inconsistency of survey data and so on. Therefore, the next step is to form profitable customers group (PCG) by removing

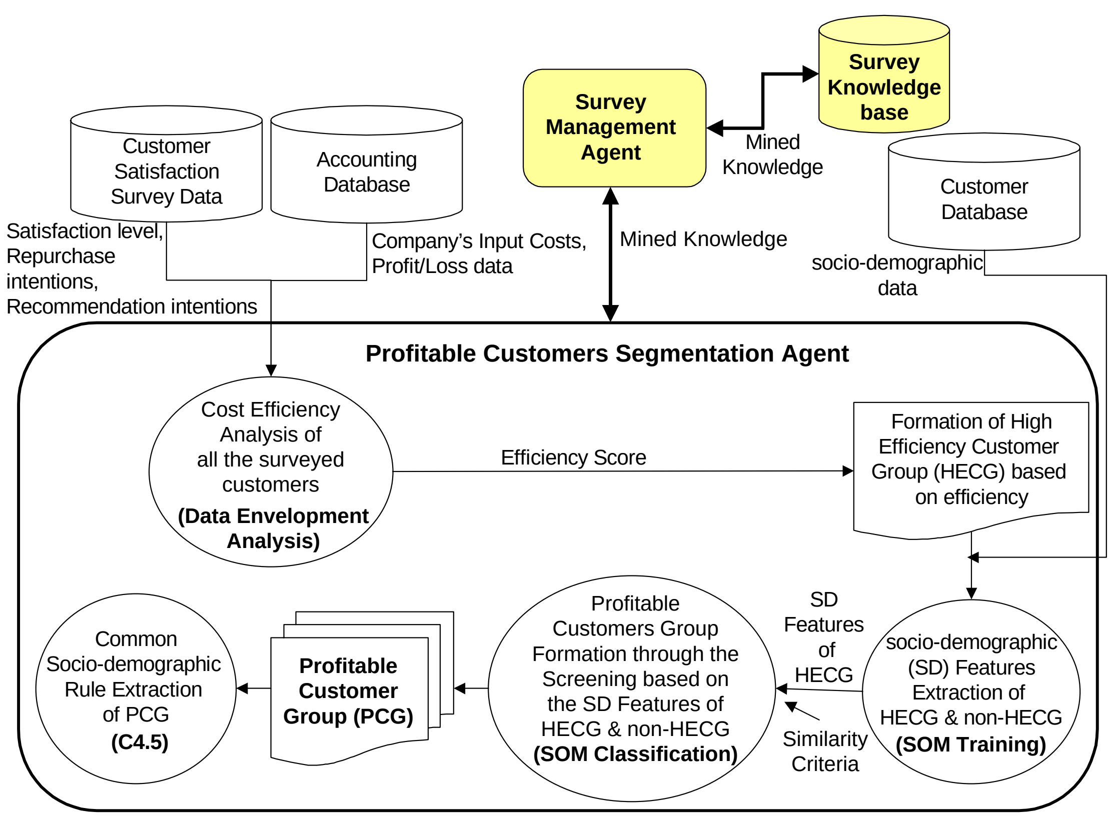

Fig. 2. Profitable customers segmentation process in PCS agent.

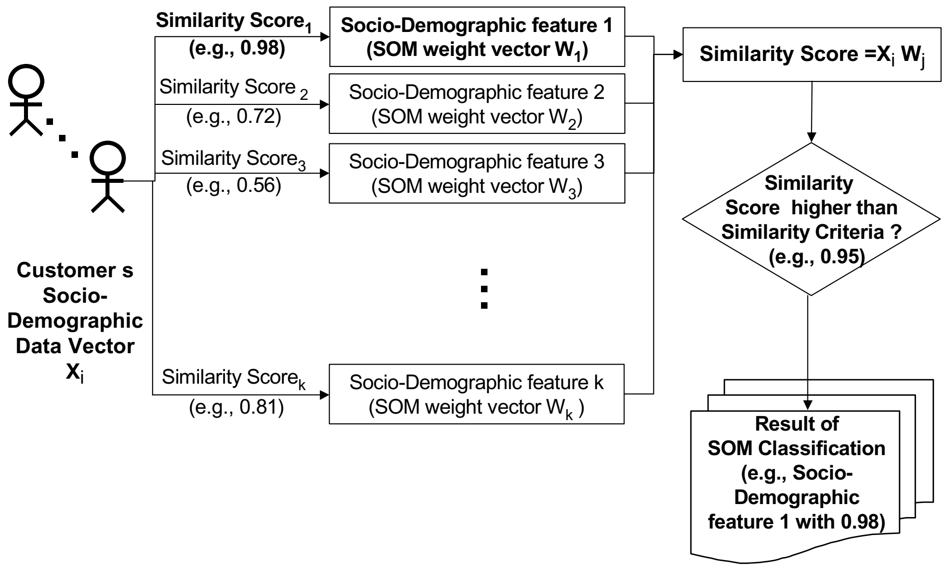

Fig. 3. SOM classification concept.

undesirable customers that have similar socio-demographic features to non-HECG's ones among the customers belonging to HECG.

PCS agent first extracts the socio-demographic (SD) features of HECG's customers and classifies them into the extracted features using SOM (Self-Organizing Map), a special type of neural network using an unsupervised learning scheme (Kohonen, 1989).

In other words, through SOM Training of HECG's SD data from customer database, PCS agent produces SOM weight vectors with the SD features information of HECG. And then through SOM Classification process as shown in Fig. 3, PCS agent classifies all the surveyed customers into the extracted SOM weight vectors (i.e. SD features) of HECG.

The SOM Classification process (Lee, You, & Park, 2001) produces *similarity scores* by using the inner product (or dot product) between the extracted SOM weight vectors (i.e. $\mathbf{W}_j$, $j=1,\ldots,k$) that show the SD feature patterns of HECG and the SD data vector of a survey customer (i.e. $\mathbf{X}_i$). The *similarity score* indicates the level of similarity between two vectors, the extracted SOM weight vector and observed SD data vector of customer belonging to the HECG.

In the case of *similarity score* higher than the *similarity criteria* that sets a lower limit on the level of similarity, the customer is classified into the corresponding SD feature of HECG. In other words, the customer has that SD feature of HECG. For example, in Fig. 3 the customer's socio-demographic data vector, $\mathbf{X}_i$, is classified into the SD feature 1 (i.e. $\mathbf{W}_1$) among k SD features because the *similarity score* of two vectors, 0.98, is higher than the predefined *similarity criteria*, 0.95.

Through the SOM classification with high *similarity criteria* setting as shown in above example, PCS agent composes the profitable customers group (PCG) by selecting the customers with the very similar SD features to HECG's ones among all the surveyed customers.

Finally, PCS agent extracts the common SD rules of customers belonging to the PCG by C4.5 (Quinlan, 1993), a decision tree learning tool, and then these mined rules are accumulated in the survey knowledge base of SPCSS to use them for customer management later.

3.2. Non-profitable customer's priority order determination by PCS agent

After PCS agent segments profitable customers, it evaluates the possibility that non-profitable customer (non-PC) is converted to profitable customer (PC) for all non-PCs and decides the priority order according to the size of possibility. Fig. 4 shows the determination process of priority order in PCS agent.

As Fig. 4 shows, PCS agent detects the discriminating factors (e.g. quality-related and/or service-related factors in the customer satisfaction survey) except socio-demographic (SD) ones that divide the survey customers into PCG and non-PCG through the analysis of survey data using C4.5. PCS agent chooses the nodes appeared in a decision tree created by C4.5 discriminating factors.

It extracts the features of those discriminating factors of PCG through SOM Training and evaluates the *similarity score* between the customers' pattern belonging to non-PCG and the extracted PCG's features in discriminating factors through SOM Classification process. Finally it decides the priority orders according to the size of *similarity score* for all non-profitable customers (non-PCs). The resulted priority orders give a company the ability to prioritize customer interactions.

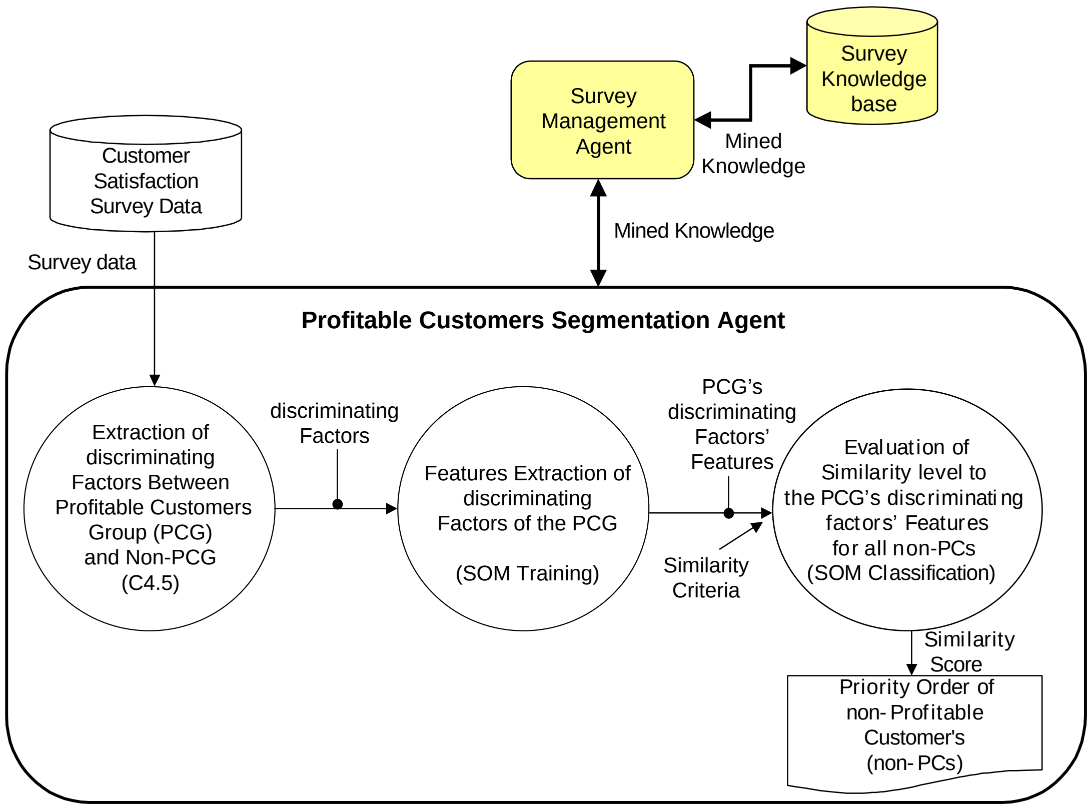

Fig. 4. Priority order determination process in PCS agent.

4. A case study on a motor company's profitable customers segmentation

We implemented a web-based SPCSS prototype and validated the effectiveness of our approach through the customer satisfaction survey data of T Motor company.

The Survey management (SM) agent in SPCSS conducted a customer satisfaction survey of T company via e-mail and 491 customers responded to this survey. SM agent selected the survey customers from the customer database of T company. All respondents were asked to rate 24 questions addressing factors such as product quality, customer service, overall satisfaction, repurchase intentions, and word-of-mouth intentions and so on.

For each question, a seven-point scale was used with the ratings of very dissatisfied (1.0), dissatisfied (2.0), somewhat dissatisfied (3.0), fair (4.0), somewhat satisfied (5.0), satisfied (6.0), very satisfied (7.0). A 'don't know' option was also provided.

4.1. T company's profitable customer segmentation by PCS agent

PCS agent executed a DEA in order to select the customers that had higher efficiencies about their output from the T company's input costs for them. The DEA estimated the relative efficiency of all 491 surveyed customers that have common 5-input (i.e. marketing cost, production cost, inventory cost, delivery cost, service cost) and 4-output (i.e. the level of overall customer satisfaction, repurchase intentions, word-of-mouth intentions obtained from the customer satisfaction survey database and their profit/loss to the T company derived from the accounting database of it).

In this study, we composed HECG (High Efficiency Customer Group) by selecting the customers with higher than 0.95 efficiency score in the result of DEA analysis (Table 1), which had high efficiency scores ranging from 0.950214 to 1.0.

Table 1
High efficiency customer group

High efficiency customer group	Total number	Customer Id.
The customers with higher than 0.95 efficiency score	88	1, 3, 4, 6, 8, 9, 11, 14, 15, 16, 18, 20, 22, 25, 27, 30, 32, 34, 35, 37, 41, 42, 43, 44, 45, 50, 51, 52, 63, 71, 73, 81, 86, 90, 94, 99, 100, 101, 108, 110, 111, 116, 121, 122, 123, 126, 127, 128, 130, 133, 137, 139, 140, 143, 146, 167, 171, 200, 211, 220, 260, 270, 271, 277, 278, 281, 288, 289, 301, 304, 309, 311, 325, 328, 333, 376, 380, 420, 421, 428, 430, 445, 447, 459, 476, 487, 492, 497

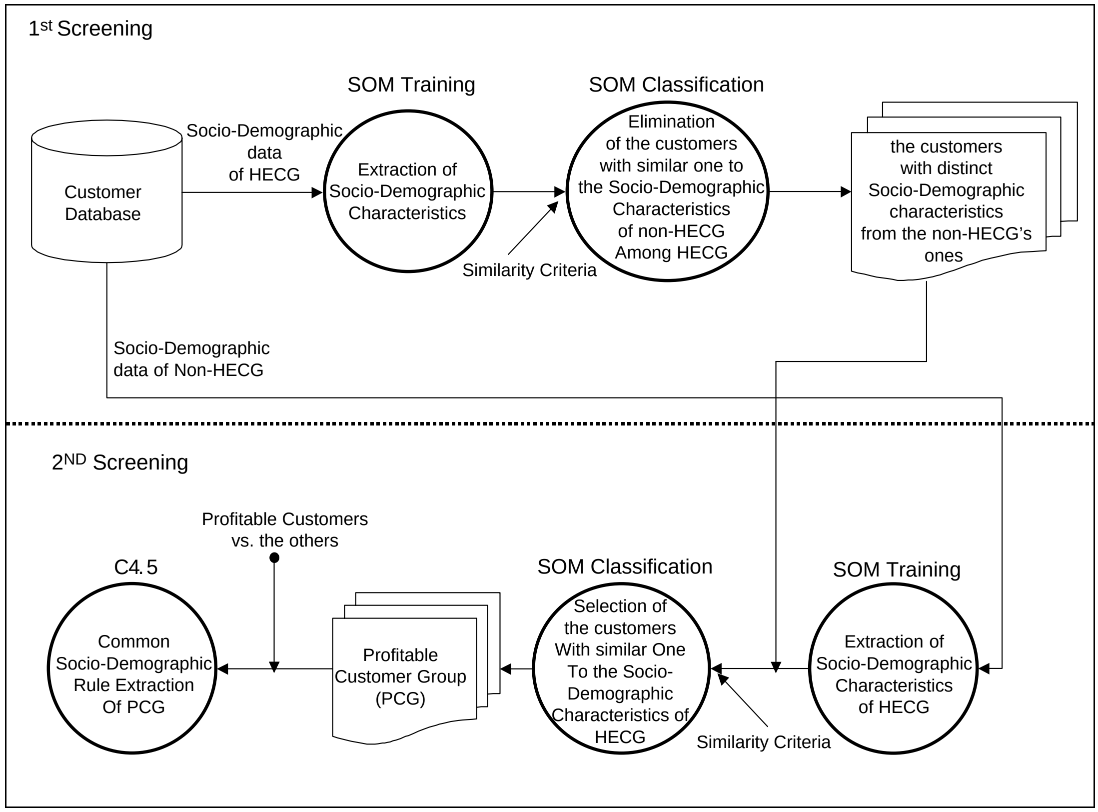

Fig. 5. Two-stage screening processes for the profitable customer segmentation.

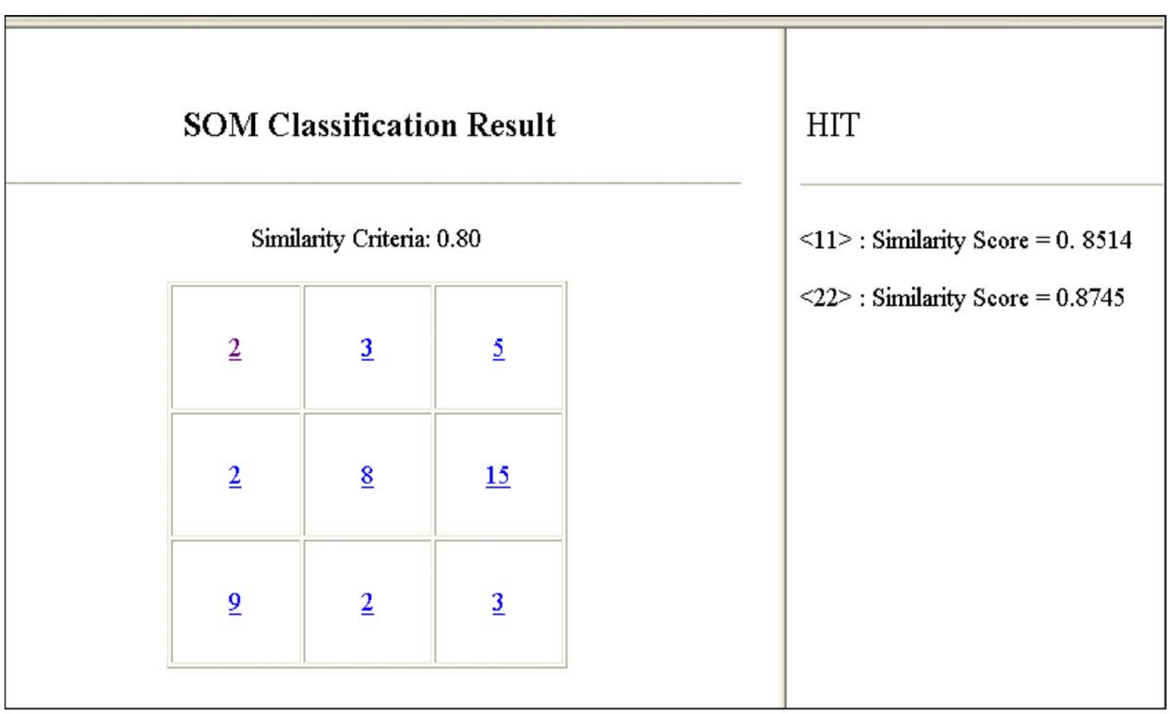

Fig. 6. SOM classification result for all 88 HECG 's customers in *similarity criteria* setting of 0.8.

Table 2
The group of customers with the distinct socio-demographic features from non-HECG's ones

Customer group	Total number	Customer Id.
The customers with the distinct socio-demographic features from non-HECG's ones	42	1, 3, 4, 6, 8, 9, 14, 15, 18, 27, 30, 32, 34, 35, 37, 41, 42, 43, 50, 63, 73, 81, 86, 94, 99, 101, 111, 116, 122, 143, 146, 271, 281, 301, 311, 333, 380, 428, 445, 459, 487, 492

PCS agent performed the SOM Training and Classification process for 88 customers belonging to HECG and all non-HECG, 403 customers. To segment profitable customers and extract its common socio-demographic rules, PCS agent performed two-stage screening processes (Fig. 5).

The first screening stage was to remove the customers who had the socio-demographic (e.g. age, income, marital status, etc.) features of non-HECG among the customers belonging to HECG. First, PCS agent extracted nine socio-demographic features of non-HECG through SOM Training of non-HECG's socio-demographic data extracted from customer database and classified all 88 customers belonging to HECG into the extracted nine socio-demographic features through SOM Classification with *similarity criteria* setting of 0.8 (Fig. 6).

In the SOM Training and Classification, PCS agent used the normalized value of extracted socio-demographic data because of different socio-demographic variables' scales such as age, income.

In Fig. 6, nine grids represent the nine socio-demographic features of non-HECG and the number in each grid indicates the number of customers belonging to that socio-demographic feature. Two customers, Customer 11 and 22, belonging to HECG were classified into the socio-demographic feature 1 of non-HECG. In other words, although they belonged to HECG, they had the similar socio-demographic feature more than 80% to non-HECG's one and thus had fairly high possibility to belong to the non-profitable customers group. Therefore, PCS agent removed them.

By removing all the customers belonging to all nine grids in Fig. 6, PCS agent formed a new customer group that had the distinct socio-demographic features from non-HECG's ones (Table 2).

The second screening stage was to remove the customers who did not have the socio-demographic features of HECG among 42 customers in Table 2. In this time, PCS agent extracted nine socio-demographic features of HECG through SOM Training of HECG's socio-demographic data and classified all 42 customers into the nine extracted socio-demographic features of non-HECG through SOM Classification with *similarity criteria* setting of 0.8.

Table 3
Profitable customers group

Customer group	Total number	Customer Id.
Profitable customers	11	1, 8, 15, 18, 30, 41, 94, 146, 271, 311, 487

In conclusion, the customers belonging to the nine socio-demographic features of HECG obtained by this SOM Classification had similar socio-demographic features to HECG's ones and had distinct socio-demographic features from non-HECG's ones at the same time. In this study, we called those customers profitable customers and Table 3 showed the customers belonging to profitable customers group (PCG).

Finally, PCS agent generated the decision tree through the use of C4.5 that showed common socio-demographic rules of 11 customers belonging to the PCG (Fig. 7).

We found three socio-demographic rules of customers belonging to PCG by interpreting the generated decision tree; Rule 1 that the sex is '1' (i.e. man), the life style is '1' (i.e. married and branch family), and the age is 43 and less, Rule 2 that the sex is '1' (i.e. man), the life style is '2' or '3' (i.e. joint living with family or singleness), and the age is 36 through 50. Rule 3 that the sex is '1' (i.e. man), the life style is '2' or '3' (i.e. joint living with family or singleness), and the age is more than 56.

The potential customer with the conformable socio-demographic pattern to these three rules should be intensively managed because his/her possibility to belong to the PCG is high.

4.2. T company's non-profitable customer's priority order determination by PCS agent

To identify discriminating factors that exactly classified the customers into PCG and non-PCG among 11 factors such as Design, Color, Ride Comfort, Engine Noise, Fuel Economy, Price, After Service, etc. which are questioned in

```
sex > 1 : non (37/1.4)
sex <= 1 :
|  sty <= 1 :
|  |  age <= 43 : PCG (4.0/1.0)
|  |  age > 43 : non (16/1.1)
|  |  sty > 1 :
|  |  age <= 35 : non (15/2.1)
|  |  age > 35 :
|  |  |  age <= 50 : PCG (3.0/1.0)
|  |  |  age > 50 :
|  |  |  |  age <= 55 : non (8.0/1.2)
|  |  |  |  age > 55 : PCG (2.0/1.0)
```

Fig. 7. The decision tree showing PCG's common socio-demographic pattern.

```
RidC  <= 5 : non
RidC  > 5 :
|       EngNo  <= 4 : non
|       EngNo  > 4 :
|         |       FuelE <= 6 : non
|         |       FuelE > 6 :
|         |         |       Desi<=5 :  non
|         |         |       Desi> 5 : PPCG
```

Fig. 8. Decision tree of PCG and non-PCG in 11 surveyed factors.

Table 4
SOM Classification result of non-PCG's customers in *similarity criteria* setting of 0.9

Customer Id	SOM classification result (four discriminating factors' feature)	*Similarity score*
3	Feature 2	0.9288
	Feature 4	0.9113
27	Feature 2	0.9075
333	Feature 4	0.9178
	Feature 2	0.9102

the T Motor company's customer satisfaction survey, PCS agent performed C4.5.

Fig. 8 showed the generated decision tree through the use of C4.5 in PCS agent.

We found that the discriminating factors for the PCG and non-PCG classification were Ride Comfort, Engine Noise, Fuel Economy, and Design from the decision tree shown in Fig. 8. In order to be PCG's customer, the satisfaction level about Ride Comfort should exceed five (i.e. Somewhat satisfied), the satisfaction level about Engine Noise exceed four (i.e. Fair), the satisfaction level about Fuel Economy exceed six (i.e. Satisfied), and the satisfaction level about Design exceed five (i.e. Somewhat satisfied).

In the next step, PCS agent extracted the nine features of PCG's customers in terms of four discriminating factors (i.e. Ride Comfort, Engine Noise, Fuel Economy, and Design) through SOM training and classified all non-PCG's customers into the extracted nine features of PCG in four discriminating factors through SOM Classification with *similarity criteria* setting of 0.9 (Table 4).

Table 4 showed that only three customers (i.e. Customer 3, 27 and 333) among non-PCG's customers could be classified. Customers 1 had higher *similarity score* as 0.9288 about the Feature 2, Customer 333 had higher *similarity score* as 0.9178 about the Feature 4, and Customer 27 had a *similarity score* as 0.9075 about the Feature 2.

Therefore, PCS agent decided the priority order of three customers in order of Customer 3, 333 and 27 according to the size of *similarity score*. We could infer that Customer 1 among them had the highest possibility to move into PCG.

5. Conclusion

To intelligently segment profitable customers of a company in terms of their profitability, we present an easy and efficient alternative approach based on the mining of customer satisfaction survey, socio-demographic and accounting database instead of using a complicated customer profitability model.

First, the presented approach uses DEA to find out the customers with higher cost efficiency, High Efficiency Customer Group (HECG), among all the surveyed ones about their output from a company's input costs for them. And then it uses SOM to form the profitable customers group (PCG) by removing undesirable customers among HECG's customers. Finally, it successively uses C4.5 and SOM to decide the priority orders of non-PCG's customers.

We also propose a survey-based profitable customers segmentation system (SPCSS) that conducts the customer satisfaction survey and those mining processes for the profitable customers segmentation.

When our work is used in practice, the appropriate setting of efficiency score criterion in DEA analysis and *similarity criteria* in the SOM Classification is required. We expect that our study will offer an opportunity to use various survey data including customer satisfaction survey actively and develop an intelligent methodology for profitable customers segmentation.

References

Anderson, E. W., & Mittal, V. (2000). Strengthening the satisfaction–profit chain. *Journal of Service Research*, *3*, 107–120.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: marketing models and applications. *Journal of Interactive Marketing*, *12*(1), 17–30.

Charnes, A., Cooper, W., Lewin, A. Y., & Seiford, L. M. (1994). *Data envelop analysis*. Dordrecht: Kluwer.

Gupta, S., & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, *17*(1), 9–24.

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, *26*(2), 181–188.

Lee, J. H., You, S. J., & Park, S. C. (2001). A new intelligent SOFM-based sampling plan for advanced process control. *Expert Systems with Application*, *20*(2), 133–151.

Griffin, J. (2003). Customer segmentation: divide and prosper, *iQ Magazine*, Cisco Systems, March/April.

Kohonen, T. (1989). *Self-organization & associative memory* (3rd ed.). Berlin: Springer.

Quinlan, R. J. (1993). *C4.5: Programs for machine learning*. San Mateo, California: Morgan Kaufmann Publishers.

Reichheld, F. F., & Frederick, F. (1996). *The loyalty effect*. Harvard Business School Press.