PERGAMON

Expert Systems with Applications 20 (2001) 187–200

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mixed-initiative synthesized learning approach for web-based CRM

Soe-Tsyr Yuan*, Wei-Lun Chang

Information Management Department, Fu-Jen University, Taipei, Taiwan

Abstract

The issue of customer relationship management has emerged rapidly. Customers have become one of the most important considerations to new companies being built. Accordingly, customer retention is a very important topic. In this paper, we present a mixed-initiative synthesized learning approach for better understanding of customers and the provision of clues for improving customer relationships based on different sources of web customer data. The approach is a combination of hierarchical automatic labeling SOM, decision tree, cross-class analysis, and human tacit experience. The objective of this approach is to hierarchically segment data sources into clusters, automatically label the features of the clusters, discover the characteristics of normal, defected and possibly defected clusters of customers, and provide clues for gaining customer retention.

Keywords: Customer relationship management; Customer retention; LabelSOM; Decision tree

1. Introduction

Customer relationship management (CRM) is one of the fastest growing business technology initiatives since the web. That growth rate will only continue to climb higher as businesses increasingly recognize that in order to achieve a sustainable competitive advantage, they must first understand and then please their customers. In other words, understanding customers and strengthening relationships with them is essential in today's demand-driven economy as customers are not just buying products; they are buying their relationship with businesses.

CRM not only discusses traditional customer problems but also provides an integrated solution to resolve internal problems in businesses, including “Customer Marketing”, “Sale”, “Customer Services”, and integrating people, process and technology to form a revolutionary concept (www.dci.com/crm; www.crm-forum.com.crm_forum_-white_papers/dcr/sld01.htm). The ultimate goals of CRM are to acquire new customers, retain old customers, and grow customer profitability.

A report from Customer Retention Practice Newsletter in 1998 pointed out “the typical company derives 80% of its profit from 20% of the customers base” (www.customer-loyalty.org). Therefore, we can understand that the role of the customers in the enterprises becomes more and more important. It also explains why the companies employ all kinds of competition and marketing techniques to retain customers. In other words, the importance of the customers to the companies is greatly increased. Consequently, how to verify customers' demands and how to retain customers are becoming the most important issues (retention.harrisblack-intl.com/solutions).

On the other hand, although all enterprises have their own customers, they still have to continuously discover potential target customer. Therefore, “Customer Loyalty” is the most important link to the potential commercial opportunities because the companies can employ existing customers to stimulate re-purchase and analyze their characteristic attributes to predict potential customer segments.

In general, there are three phases in the CRM Lifecycle:

- Integration: output centralized customer data from difference sources.
- Analysis: provide a deeper understanding of customer behavior and needs.
- Action: provide positive impact on customer relationships.

Web-based CRM means the sources of customer data are generated from the customer-Web interactions, and its lifecycle also follows the three phases mentioned above.

Currently, there are a lot of CRM solution providers (retention.harrisblackintl.com/solutions), which tackle the CRM problems based mainly on statistical approaches, which mostly assume linearity between data, and delicate human consultancy, accordingly requiring much time and effort. This paper presents an alternative novel synthesized

* Corresponding author. Tel.: +886-2-2369-3220; fax: +886-2-2369-3220.

E-mail address: yuan@tpts1.seed.net.tw (S.-T. Yuan).

0957-4174/01/$ - see front matter
PII: S0957-4174(00)00058-0

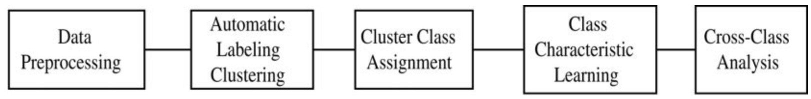


Fig. 1. The conceptual framework of the mixed-initiative synthesized learning approach.

approach for a better understanding of customers and the provision of clues for improving customer relationships based on different sources of web customer data. This approach differs from previous ones in the following two ways: (1) visual and nonlinear data analysis, (2) fully integrated mixed-initiative customer understanding and relationship upgrading.Visual data analysis renders the task of data analysis more friendly and understandable, while nonlinear data analysis makes it possible to uncover complex relationship embedded in data. A fully integrated mixed-initiative approach makes it easier for users to collaborate and therefore come up with good results.

Our approach is a synthesized learning approach, which is a combination of hierarchical automatic labeling SOM (H-LabelSOM) and decision tree.

The rest of the paper is organized into four sections. Section 2 describes the framework of the approach. Section 3 provides the descriptions of the architecture and the components of the approach, automatic-labeling SOM (LabelSOM) (Andreas, 1997, 1999a,b), decision tree, and cross class analysis, and the way we synthesize these components. An example is also given to show how the approach gains better understanding of customers and provides clues for improving customer relationships. Section 4 provides evaluation results of the approach. Finally, a discussion and a conclusion are made in Section 5.

2. The conceptual framework of the mixed-initiative synthesized learning approach

In this section, we describe the conceptual framework, the architecture, and the properties of the mixed-initiative synthesized learning approach.

The conceptual framework shown in Fig. 1 shows that there are five major tasks associated with the mixed-initiative synthesized learning approach:

1. Web-based customer data is integrated for analysis.
2. The centralized customer data is clustered by the H-LabelSOM that generates visual diagrams showing all clusters with labeled characteristic features.
3. Human analysts assign classes (Normal, Defected, Possibly Defected) to clusters based on the labeled features of clusters.
4. The clusters of data with class labels serve as a hold-out sample data for the decision tree, which inductively inference the characteristic rules for the customers in Normal/Defected/Possibly Defected classes.
5. Obtain valuable knowledge for the problems of CRM from the cross-class analysis of rules generated at the fourth step.

The properties exhibited by the conceptual framework are five-fold:

1. Web-based CRM: We used three sources of customer data generated from the customer-Web interactions of the Taiwan branch of the worldwide leading printer company:
   - Member.txt: containing the basic information and profile of customer members; 31,298 records, each of which has 23 attributes such as member type (*memtype*), birth date (*bdate*), sex (*sex*), occupation (*wtype*), living area (*area*) etc.
   - product.txt: containing members' purchasing history; 6441 records, each of which has 11 attributes such as product type (*ptype*) etc.
   - visitlog.txt: containing members' behavior at the company's Web site; 20,286 records, each of which has eight attributes such as the first visit time (*fvisit*), last visit time (*lvisit*), the count of visits between *fvisit* and *lvisit* (*viscnt*), and whether or not the customer subscribes to e-news (*orden*) etc.

   With database operations, we clean and integrate the different sources of data into one centralized database containing 3658 records of registered customer member data, each of which contains 40 attributes as shown in Table 1, by removing duplicate data or blank data, integrating the data with the aid of the primary key, and converting categorical attributes into numerical attributes. We selected 11 useful attributes for the subsequent phase of analysis. These are *memtype*, *bdate*, *sex*, *area*, *vistcnt*, *wtype*, *firstvisit*, *lastvisit*, *orden*, *ptype*, and *visittype*. The remaining 29 attributes, such as email address, the e-mail server, the serial number of product purchased *etc.,*are still accessible for global view of the data after analysis. All of the attribute values of the records will be normalized into numbers ranging between 0 and 1 before feeding into the next phase of visual clustering analysis.
2. Visualization clustering analysis method with automatic labeling: In general, businesses have a huge amount of customer data. Therefore, it is important to transform the huge customer data into useful information. Our approach utilizes the visualization clustering analysis method, LabelSOM, to generate visual clusters of customers with labeled features for easier understanding of the segmentation of customers. The labeled features of a

Table 1
Integrated attribute names and their meaning

Attribute name	Meaning
memtype	Member type (1, Individual member; 2, Company member
sex	Member sex (0, Male; 1, Female; 2, Company)
bdate	Member age
area	Member living area (1 ~ 25)
	1.台北市 2.台北縣 3.基隆市 4.宜蘭縣 5.花蓮縣 6.桃園縣 7.新竹市 8.新竹縣 9.苗栗縣 10.台中市 11.台中縣 12.彰化縣 13.南投縣 14.嘉義縣 15.嘉義市 16.雲林縣 17.台南市 18.台南縣 19.台東縣 20.高雄市 21.高雄縣 22.屏東縣 23 金門縣 24.連江縣 25.澎湖縣
vistcnt	the count of total visits
wtype	Member occupation type (memtype = 1, individual occupation type; memtype = 2, business type)
ptype	Product purchased (5 categories, 22 items)
	1.DeskJet 400C, DeskJet 420C, DeskJet670C, DeskJet695C, DeskJet 710C, DeskJet720C, DeskJet 890C, DeskJet 895Cxi, Desk Jet 1120C 9 items
	2.ScanJet4100C, ScanJet5100C, ScanJet6200C, ScanJet6250C, PhotoScanner 5 items
	3.OfficeJet 635, OfficeJet 1150C, OfficeJet 1170C, OfficeJet 1175C 4 items
	4.Digital Camera C20 1 item
	5.CDRW 7200I, CDRW 7200E, CDRW 8100I 3 item
firstvisit	First visit time
lastvisit	Last visit time
orden	If or not subscribe e-news
vsittype	Indicator if the e-mail software of the member can open html documents

cluster highlight the most important attributes characterizing the cluster, which were unattainable in previous CRM approaches. Furthermore, LabelSOM is a nonlinear clustering method, which can uncover nonlinear cluster boundary relationships, which is also unattainable in most of the previous statistical analysis approaches.

3. Mixed-initiative: This paper presumes that usually with their tacit experience, human analysts know the class categories (Normal, Defected, Possibly Defected) of a given cluster based on a few highlighted features/values of the cluster, but it is hard for them to come up with a set of rules that delineate the necessary conditions of different class categories.
4. Customer class characteristics rules are generated with a decision tree: The labeled features of clusters with class labels enables a decision tree to inductively inference the characteristic rules of customers in Normal/Defected/Possibly Defected classes. Subsequent comparison between cross-class characteristics can then be made for providing clues to improve customer relationships.
5. Human-computer interaction: Human analysts play a vital role in the CRM process. In our approach, users have the ability to set a threshold for constraining the size of clusters and for making selections of attributes for subsequent analyses as well as the task of cluster class assignment. Section 4 will evaluate the performance of our approach with different settings of these dependents controlled by users.

3. The synthesized learning approach

The synthesized learning approach we present is a combination of two methods: LabelSOM (Andreas, 1997, 1999a) and decision tree (yoda.cis.temple.edu:8080/UGAIWWW/lectures/C45; www.rulequest.com/see5-info.html). In this section, we describe these two methods as well as the way we synthesize these methods. At the end, an example is given to show how the approach provides solutions for CRM problems.

3.1. LabelSOM

LabelSOM itself is a combination of the visualization clustering method SOM (Kohonen, 1997) and the technique of automatic labeling.

3.2. SOM

The visualization clustering method SOM was proposed by Kohonen and has recently became one of the most popular clustering methods. It is not only applied in resolving engineering problems but also employed to analyze data (Kaski & Kohonen, 1996; Back, Sere, & Vanharanta, 1997; Goser, 1997; Vesanto, 1997). Its strength is to project high dimension data into two-dimensional grids while preserving its topology.

Assume each data record has many attributes called dimensions. The main ideas of the visualization clustering

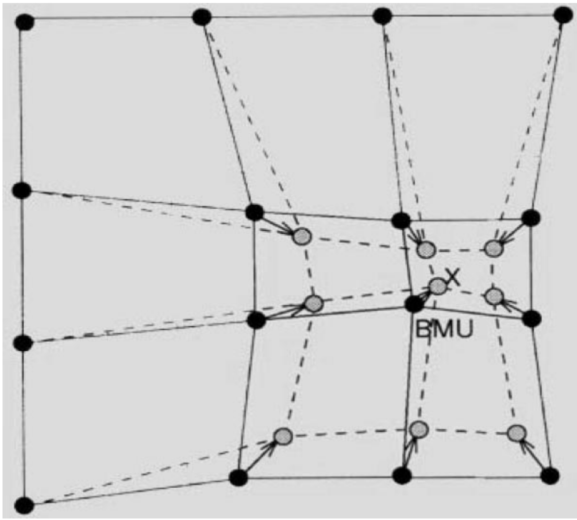


Fig. 2. "Competitive Learning" Concept in the SOM.

method SOM are as follows:

- If the data points are close to each other in N-dimension data space, they should be close each other as well in two-dimension projected space.
- This method employs a "competition" concept: all data points are input vectors and all neurons in two-dimensional space are output (also called competitors); each neuron is expressed by a model vector with the same dimensions as data points; all model vectors are compared with an input vector; the closest neuron is the winner (best matching unit) and the model vectors of itself and its neighboring neurons are adjusted automatically in order to make them closer to the input vector as shown in Fig. 2. Below is the formula for adjustion:

  $m_i(t+1) = m_i(t) + \alpha(t)[x(t)\_m_i(t)]$

  for each $i \in N_c(t)$

  where t is time, $N_c(t)$ is the neighborhood kernel function that identifies the set of neurons near to a best matching unit c, and $\alpha(t)$ is the learning rate between 0 and 1.
- After the learning process, each neuron represents a set of data points. Data points that are near each other in the input space are mapped to nearby map neurons in SOM.

There exist two common functions for $N_c(t)$: bubble function and gaussian function. The most applied one is gaussian function. The reason for adjusting neighbor nodes is that it is necessary to make a whole cluster represented by neighboring neurons get closer to the data they represent in order to maintain the neighborhood relationship.

The major difficulty in dealing with SOM maps obtained after learning is that *human visual decisions* on the nonlinear boundaries between clusters and the number of clusters are unavoidably required.

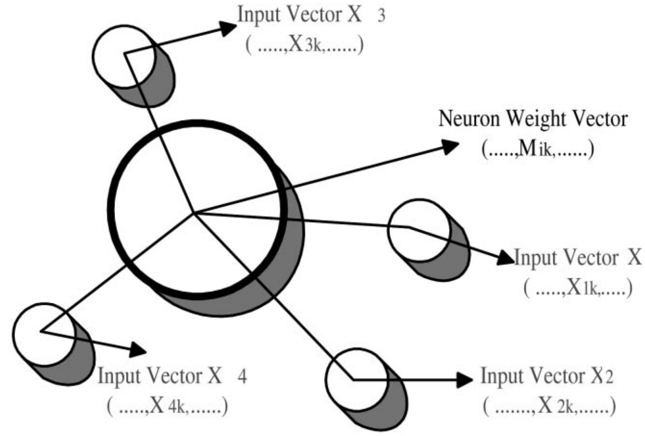


Fig. 3. The concept of the distance measure of the automatic labeling technique.

3.3. Automatic labeling

Since Kohonen proposed SOM concept in 1989, many researchers have been trying to improve SOM and develop new concepts. For example, even though SOM has the visualization capability, it still cannot automatically label the features of clusters generated. Andreas (1999a) proposed a novel approach, LabelSOM, which automatically labels the features of clusters generated by SOM.

Automatic labeling aims to automatically filter the huge amounts of customer data records in a cluster into features. By labeling, we mean to output the important attributes that contribute to the formation of the cluster. For instance, with the automatic labeling approach, only five out of the original 11 attributes are considered important and serve as the features of the cluster.

The way of determining important attributes is through a distance measure, with which the attributes can be prioritized. Subsequently, we can take features with important priorities as the characteristic of the cluster. Below is the formula of the distance measure, and Fig. 3 shows its concept.

$$q_{ik} = \sum_{x_j \in C_i} \sqrt{(m_{ik} - x_{jk})^2}, \quad k = 1...n$$

where *qik* is the distance measure for the *k*th attribute in a given neuron *i*, *Ci* is the set of data records *xj* projecting into a same neuron, m_{ik} is the *k*th attribute value of the neuron model vector and *xjk* is the *k*th attribute value of the *xj* input vector.

In a given cluster, for each attribute, sum the distance between the attribute value of a data record in the cluster and the attribute value of the neuron model vector. If the calculated sum is smaller than a threshold, it means the attribute plays an important role for the formation of the cluster, and can be served as a feature of the cluster. On the other hand, we also can set a bound for the number of attributes to be considered. For example, if the bound is 5, the five most important attributes will be selected as the features of the cluster.

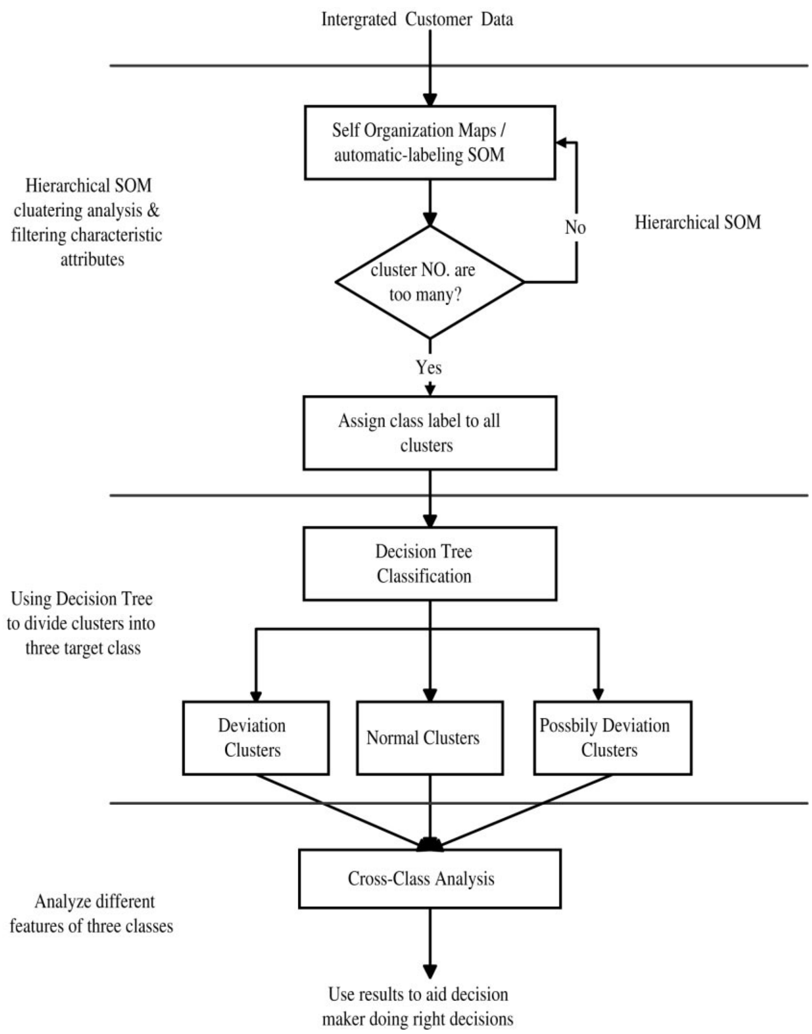


Fig. 4. The architecture of the mixed-initiative synthesized learning approach.

3.4. Decision tree

Decision tree is one of the most frequently used supervise-learning methods. It induces concepts from training examples. It represents concepts as decision trees, which also act as the classifying rules for classifying new examples. However, the problem should be considered: if irrelevant attributes are embedded in training examples, the accuracy of learned classifying rules is under question. Therefore, C5 (www.rulequest.com/see5-info.html) is the method we employ because it takes care of the problem of irrelevant attribute.

3.5. The way to synthesize the learning methods

The way we synthesize the learning methods is four-fold (as shown in Fig. 4, the architecture of the synthesized learning method):

1. Hierarchical concept + LabelSOM: We apply the hierarchical concept to LabelSOM approach and form the hierarchical automatic labeling SOM method (H-Label-Som). The objective is to automatically reduce the size of the clusters constructed out of a set of data records until a bound which can be set by users is satisfied.
   The way the hierarchical concept is employed in LabelSOM is as follows:
   - Level 0 data = the raw integrated data of records with the attributes selected by users at the beginning.
   - Level 1 data = the set of records, each of which is comprised of the labeled attributes/values generated by applying LabelSOM to the Level 0 data.
   - Level $i + 1$ data = the set of records, each of which is comprised of the labeled attributes/values generated by applying LabelSOM to data records obtained in Level i.
2. Automatic display of boundaries between clusters: Without the human visual decision on the boundaries of clusters and the number of clusters for a LabelSOM map, our approach scans the map acting like human vision and automatically visually displays the final set of distinct clusters. In the upper-left corner of Fig. 5 is a neuron map, on which

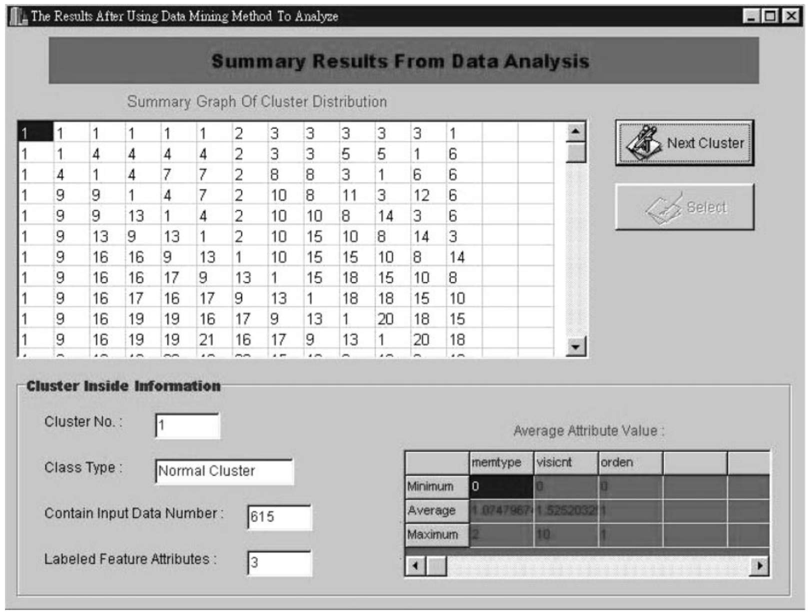


Fig. 5. The upper-left window shows the boundaries of clusters and the number of clusters; the bottom-right window shows the ranges (Maximum, Minimum, Average) for each labeled feature of a particular cluster.

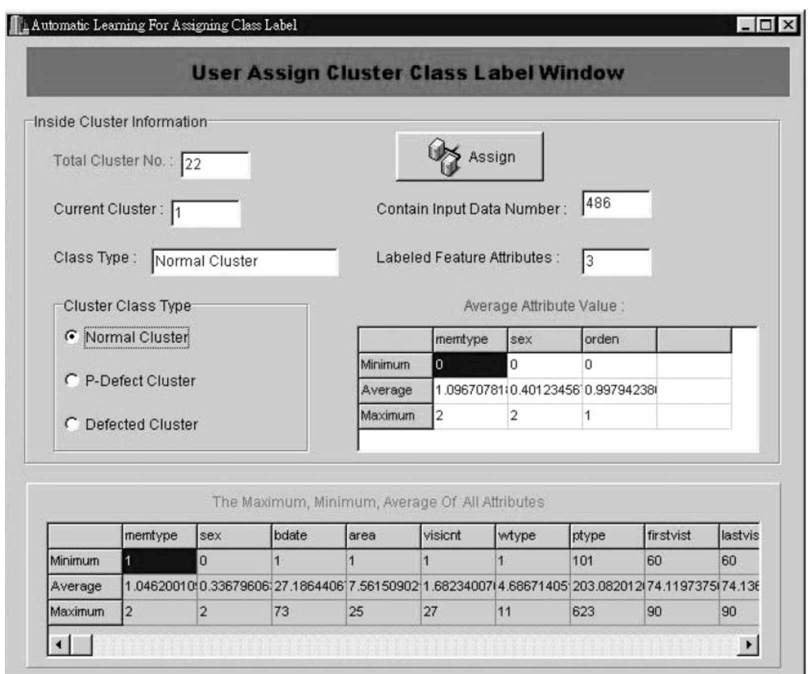


Fig. 6. The interface for cluster class assignment.

neurons in the same cluster are expressed with the same number.

The way the approach scans a LabelSOM map for determining the set of clusters is as follows:

- Starting from the upper-left neuron taken as a cluster containing only one neuron, a neighbor neuron is assigned to the same cluster as the neuron only when the distance between them is below a threshold, and is marked as a checked neuron. This process repeats until the boundary of the cluster is settled.
- Each unchecked neuron is a candidate for the next phase of scanning until all neurons of the map are checked.

3. Mixed-initiative cluster-class assignment: The mixed-initiative task of cluster class assignment bridges the unsupervised learning method, LabelSOM, and supervised learning method, decision tree. Fig. 6 shows such a

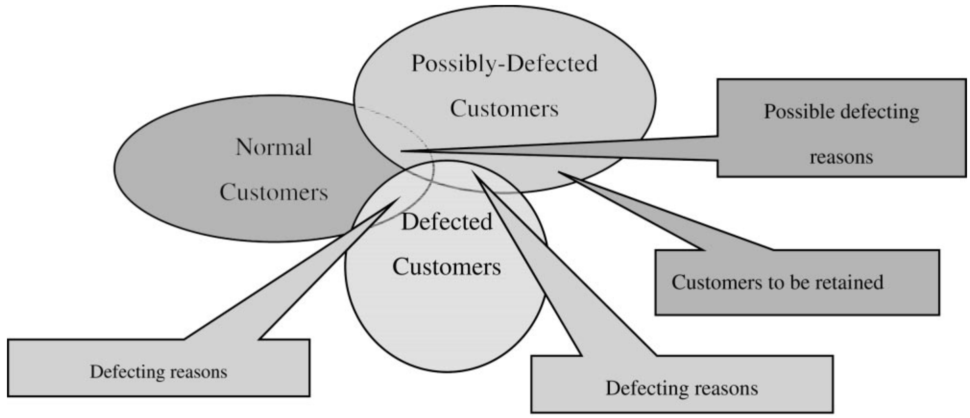


Fig. 7. The conceptual diagram of cross-class analysis.

mixed-initiative interface. For a given cluster, the sub-window on the right provides the attribute value ranges (Minimum, Average, and Maximum) of the labeled features obtained with LabelSOM. The sub-window on the bottom provides auxiliary information about the value ranges of all of the attributes in the whole data set. Based on this information provided and the tacit experience on a user's part, the user is able to assign the class for the cluster.

4. Mixed-initiative cross-class analysis: The mixed-initiative task of cross-class analysis enables the generation of valuable knowledge from data. This task is comprised of two steps:
   - As shown in Fig. 7, with the intersection operation, the approach finds the intersected attributes/values within the classes Normal, Defect, Possibly Defected respectively, followed by the findings of intersected attributes/values between any two pair of classes.
   - Comparing these six intersections, users are able to reason out the effective clues for improving the customer relationships. An example will be shown in Section 3.7.

3.6. Metrics

This section describes three metrics used for monitoring the status of analysis and measuring the quality of results obtained:

1. Cluster Number: Users can decide their favorite number of clusters, or the number of clusters can be determined by the system through a threshold described in Section 3.5 and Section 4.2. CRMiner can reduce the size of clusters for the efficient generation of analyzed results.
2. Attribute Differential Ratio: For any two classes, an attribute differential ratio is defined as: the ratio of the number of different attributes in values between two classes over the total number of attributes under analysis.

The formula is as follows:

Attribute Differential Ratio for class I, $J = D/\mathrm{T}$,

I, $J \in \{$Normal class, Defected class, Possibly Defected class$\}$

D: the number of different attributes in values between the two classes I, J

T: the total number of attributes under analysis

A higher ratio implies it is worthy of subsequent cross-class analysis because more difference between classes can lead to more knowledge for the understanding of customers' defection. Therefore, there are three attribute-differential ratios, respectively for (Normal, Defected), (Normal, Possibly Defected), and (Defected, Possibly Defected). Subsequently, we can compute the average attribute-differential ratio that can be used to express the quality of analysis results in our approach.

Retention Rate: The retention rate is computed right after cluster class assignment. The formula is as follows:

Retention_Rate = (possibly-defected + defected) customer/total customers

If (Retention_Rate $<$ default) then give-up cross-class analysis

Fig. 8. CRMiner sysem.

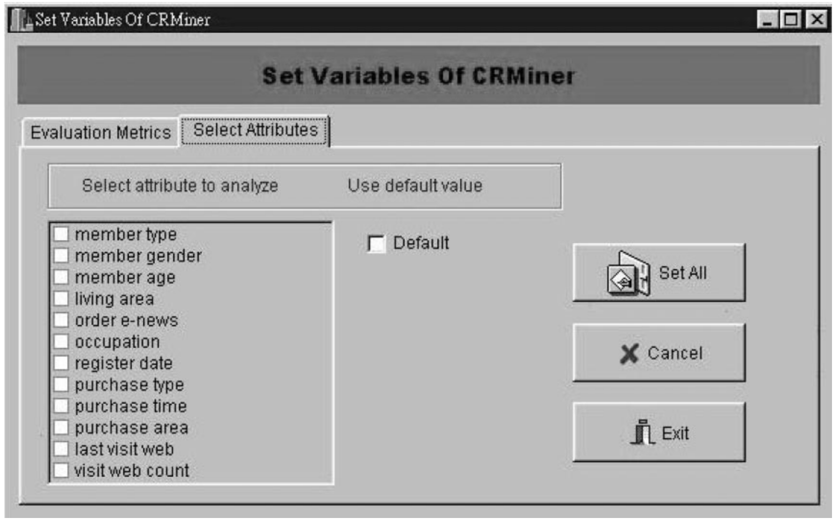


Fig. 9. The interface for users' attribute selection.

We will set the default value to 3% (the most common used in the companies (www.customerloyalty.org)) or users can decide their favorite value. If the retention rate is lower than the default value (3%), CRMiner will generally ignore those customers, but if it is higher than 3%, CRMiner will do further cross-class analysis to obtain clues of directions of customer-relationship improvement.

3.7. An example

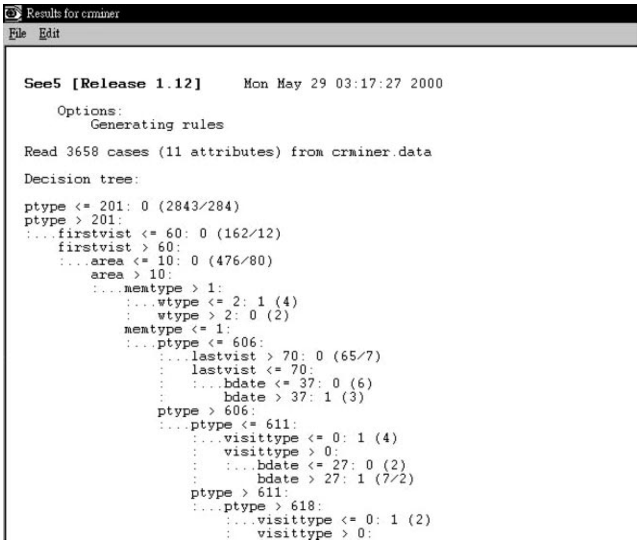
A user chooses the attributes, for instance all of the eleven attributes, to be analyzed by our system CRMiner (shown in Figs. 8 and 9). The rum-time mixed-initiative results generated have four parts (presuming the retention ratio is above the default threshold): (1) the results of unsupervised learning of H-LabelSOM (shown in the upper-left portion of Fig. 5), (2) the results of supervised learning of decision tree (shown in Fig. 10), (3) the results of the cross analysis over the characteristics of Normal, Defected and Possibly Defected classes of customers (shown in Fig. 11 and Table 2), (4) clues for gaining customer retention and acquiring directions of customer-relationship improvement.

The misclassification of this decision tree learning is

Table 2
The final tabulated cross-class analysis result with $\alpha = 9$

(a) Threshold	MAX + MIN/9
MAX	0.1761336
MIN	7.65E-10
Cluster number	20
Hierarchical cluster number	7
Rule number	42
Misclassification ratio	5.60%
Averaged attribute differential ratio	33.33%

(b) Class	Boundary	memtype	sex	bdate	area	Vistcnt	wtype	ptype	firstvisit	lastvisit	orden	Visittype
Normal	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	73	25	14	11	622	80	90	1	1
P-D	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	54	22	7	11	620	80	80	1	1
Defected	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	44	22	7	11	620	80	80	1	1
N-D	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	44	22	7	11	620	80	80	1	1
N-PD	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	54	22	7	11	620	80	80	1	1
PD-D	Lower	0	0	0	0	0	0	0	0	0	0	0
Cluster	Upper	2	2	44	22	7	11	620	80	80	1	1



```
Results for crminer
File Edit

See5 [Release 1.12]     Mon May 29 03:17:27 2000

    Options:
        Generating rules

Read 3658 cases (11 attributes) from crminer.data

Decision tree:

ptype <= 201: 0 (2843/284)
ptype > 201:
:...firstvist <= 60: 0 (162/12)
    firstvist > 60:
    :...area <= 10: 0 (476/80)
        area > 10:
        :...memtype > 1:
            :...wtype <= 2: 1 (4)
            :   wtype > 2: 0 (2)
            memtype <= 1:
            :...ptype <= 606:
                :...lastvist > 70: 0 (65/7)
                :   lastvist <= 70:
                :   :...bdate <= 37: 0 (6)
                :       bdate > 37: 1 (3)
                ptype > 606:
                :...ptype <= 611:
                    :...visittype <= 0: 1 (4)
                    :   visittype > 0:
                    :   :...bdate <= 27: 0 (2)
                    :       bdate > 27: 1 (7/2)
                    ptype > 611:
                    :...ptype > 618:
                        :...visittype <= 0: 1 (2)
                        :   visittype > 0:
```

Fig. 10. The results of decision tree learning in terms of a set of class characteristic rules.

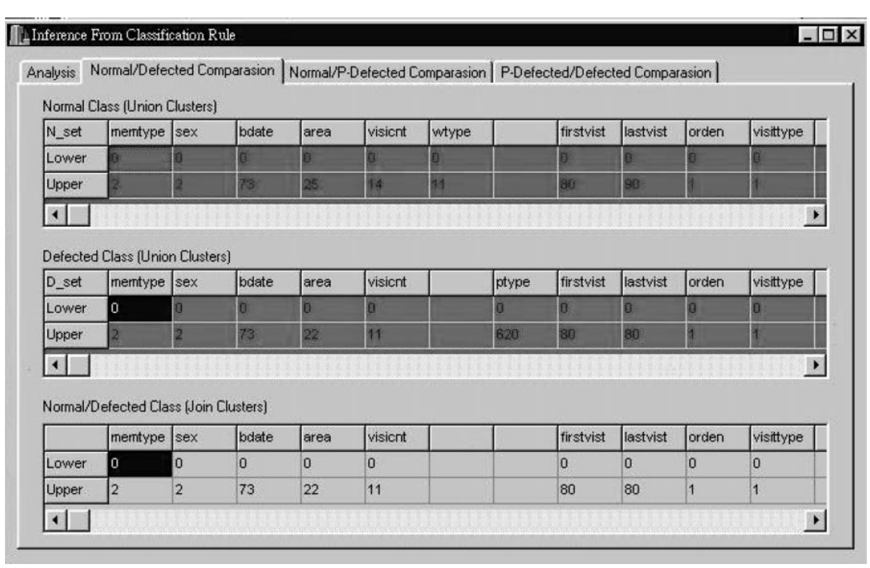
5.6%, that is, the set of learned rules misclassifies the hold-out training data at the rate of 5.6%. These results are learned from the 20 clusters of customer data constructed by the H-LabelSOM. The averaged attribute differential ratio is 33.33% (The attribute differential ratios for two classes are 5/11 = 45.4%, 5/11 = 45.4%, and 1/11 = 9%, respectively), which indicates it is worthy of doing the further analysis.

Table 2, the results of cross-class analysis, gives the company the clues about the characteristics of possibly defected customers who need customer-relationship improvement in the near future:

- Members who live in north/middle Taiwan: members in the N class live all around Taiwan (0 < = area < = 25), members in D class live above south Taiwan (0 < = area < = 22), and the intersected values for the attribute "area" of N/D, N/P-D are both associated with members live above south Taiwan (0 < = area < = 22).
- Members aged below 44: for the attribute bdate, the intersected values of P-D/D is between 0 and 44, and the intersected values of N/D and N/P-D contain those of P-D/D
- Members who used the Web site infrequently and not recently: for the attribute visitcnt, the members in the N class access the Web site up to 14 times, but the intersected values of P-D/D, N/D, and N/P-D are less than seven times. This means members who did not frequently access the Web site are the defected members. Furthermore, there are nonempty intersected values for the attribute visitcnt of N/P-D, implying the more often members visit the Web site, the less chance members become defected. In addition, for the attribute lastvisit, comparing the intersected values of P-D/D, N/D, and N/P-D with the values of N, there is 1 month difference between their last visit time.
- Members who bought the CDRW 7200I product: for the attribute ptype, the members in N class mostly bought the CDRW 8100I product, and member in D and P-D classes bought the CDRW 7200I product.

With the above characteristics of possibly defected customers, it is naturally to come to the following measurements of customer-relationship improvements, which should be exerted in the near future:

- Value the customers in southern Taiwan by providing them great service and speedy response thus increasing their loyalty.
- Special promotion toward the target market of young customers.



Inference From Classification Rule

Analysis | Normal/Defected Comparasion | Normal/P-Defected Comparasion | P-Defected/Defected Comparasion

Normal Class (Union Clusters)

N_set	memtype	sex	bdate	area	visicnt	wtype		firstvist	lastvist	orden	visittype
Lower	0	0	0	0	0	0		0	0	0	0
Upper	2	2	73	25	14	11		80	90	1	1

Defected Class (Union Clusters)

D_set	memtype	sex	bdate	area	visicnt		ptype	firstvist	lastvist	orden	visittype
Lower	0	0	0	0	0		0	0	0	0	0
Upper	2	2	73	22	11		620	80	80	1	1

Normal/Defected Class (Join Clusters)

	memtype	sex	bdate	area	visicnt			firstvist	lastvist	orden	visittype
Lower	0	0	0	0	0			0	0	0	0
Upper	2	2	73	22	11			80	80	1	1

Fig. 11. The screen dump showing a portion of the results obtained after cross-class analysis.

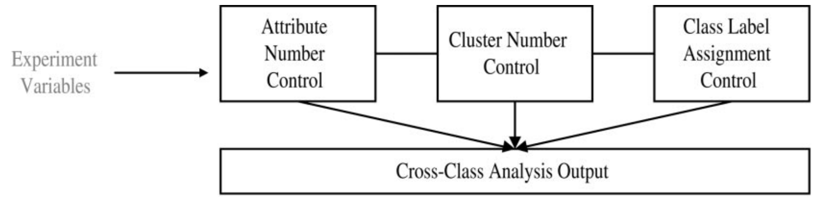


Fig. 12. Three different factors that can be controlled by users.

- Raise customers' desire of visiting and staying at the Web site by advancing the services, the information, and the functions of the Web site.
- Examine the qualities of the CDRW products and survey user' opinions on the products for the improvement of the products in the future.

4. Evaluation results

With the integrated data described in Section 2, this section examines the strengths and the limits of our approach in terms of the results obtained from various settings controlled by three different factors shown in Fig. 12 and explained as follows:

Selected attributes/attribute number: With our approach, different choice of attributes/attribute number decided by users might result in different final cross-analysis results. Thereafter, there should be a sensitivity analysis on attributes/attribute numbers in order to gain more understanding of the limits of the approach.

Cluster number: One of the most difficult problems for users to deal with in clustering analysis methods is the decision of the cluster numbers. LabelSOM renders this responsibility to human's visual capability. Our approach saves human's visual efforts by scanning the LabelSOM maps, and automatically visually displays the clusters. However, there should be the ability for users to control the size of the set of clusters. The control our approach grants is through a threshold. With different settings of the threshold, we can gain more understanding if cluster numbers influence the cross-analysis results.

Cluster class label assignment: Due to the nature of LabelSOM (random initialization of the model vectors of neurons at the beginning), the clusters generated at different times of analysis might not be exactly the same even with the same data. Subsequently, the set of clusters human analysts use for performing the task of cluster class label assignment varies each time. Therefore, there should be inspections about the stability of our approach when the tacit experience of human analysts remains stable.

4.1. The factor of selected attributes/attribute number

Considering the impact of the factor of selected attributes/attribute number, there are several different experiments with different selected attributes that must be conducted.

In order to make reasonable comparisons between experiments, we set as the fundamental attributes the eight attributes (*memtype*, *bdate*, *sex*, *area*, *vistcnt*, *wtype*, *firstvisit*, *lastvisit*). The remaining three attributes (*orden*, *ptype*, *visittype*) are open for users' selection. Therefore, the set of experiments needed to be conducted are the experiments with the following different settings, each of which is a possible user selection of attributes/attribute number: (1) the eight fundamental attributes; (2) night attributes; the one other than the fundamental attributes is selected from the set (*orden*, *ptype*, *visittype*); and (3) 10 attributes; the two other than the fundamental attributes are selected from

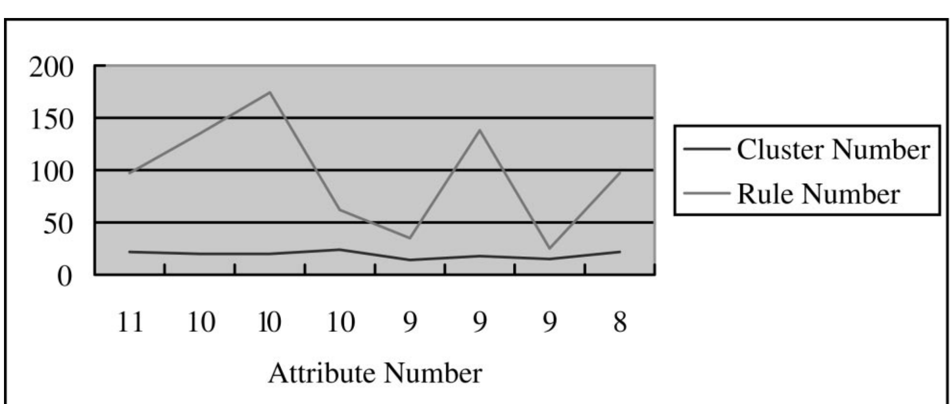


Fig. 13. The relationship between the attribute number and the cluster number/rule number in the eight experiments.

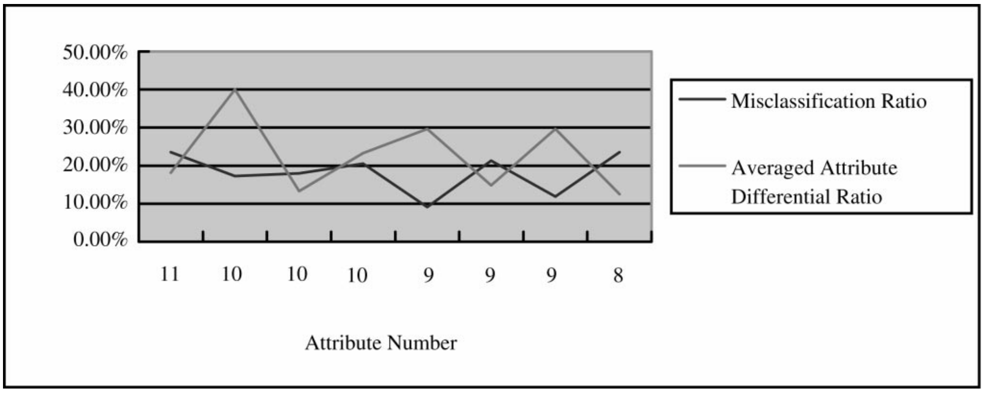


Fig. 14. The relationship between the attribute number and averaged attribute differential ration in the eight experiments.

any pair of the set (*orden*, *ptype*, *visittype*). In total, there are eight such experiments conducted.

The experiment results are shown in Table 3, Figs. 13 and 14, and a set of observations are acquired as follows:

- From Table 3, maximum/minimum number of attributes (11/8) selected does not necessarily yield the best result (highest averaged attribute differential ratio).
- From Table 3, one setting with nine attributes and the other setting with 10 attributes shown below yield better results (higher averaged attribute differential ratios):
  —eight fundamental attributes + *ptype*… (nine attributes) 40%
  —eight fundamental attributes + *ptype* + *orden*… (10 attributes) 29.63%
  These results naturally lead to an understanding of the importance of the attributes *ptype* and *orden*.
- From Fig. 13, the number of clusters does not vary much with respect to the number of attributes.
- From Fig. 14, as long as the misclassification ratio is not too high, the results are acceptable (the averaged attribute differential ratios are acceptable).

From the above observations, an understanding of the strength/limit of our approach can be acquired:

- As long as the fundamental attributes are selected appropriately, the results obtained are acceptable when the misclassification ratio is not bad. In other words, the approach maintains the stability of quality of results at the first place.
- If the importance of an attribute is represented with a numerical weight *W* ranging between 0 and 1, (*W1*A1, *W2*A2, …, *Wn*An), then binary weights are presumed in this set of experiments because an attribute is either selected or not selected for analysis. Experiments with non-binary weights may yield much better results and is worthy of future further inspection.

4.2. The factor of cluster number

With the same data set, the number of clusters is influenced by a preset threshold, which determines the extent of neighbor neurons being assigned into same clusters.

Our approach employs a simple heuristic, averaging the extremes, to express the threshold as shown in the formula below:

$$\frac{\text{Max. Attribute Value} + \text{Min. Attribute Value}}{\alpha}$$

where Max. Attribute Value is the largest normalized attribute value among among all model vectors of all neurons, Min. Attribute Value is the smallest normalized attribute value among all model vectors of all neurons, and α is the integer.

Accordingly, the impact of the factor of the cluster number can be represented by the impact of different values of α. Table 4, Figs. 15 and 16 show the experiment results using different α. Based on this, we obtain the following observations:

- From Fig. 15, the number of clusters does not vary abruptly with respect to different α.
- We get the best result when α equals 5.
- From Fig. 16, as long as the misclassification ratio is not too high, the results with different α are not bad.

From the observations, we know our approach is able to maintain the stability of quality of results at the second place, and it is best to preset α to be 5 at the beginning for good performance.

4.3. The factor of cluster class label assignment

In order to inspect the stability of our approach when the tacit experience of a human analyst remains stable, that is, when cluster class label assignment on the user's part is done consistently. Ten experiments were conducted with the same human analyst.

Table 3
The tabulated results for different settings of selected attributes/attribute number

									Average value
Attribute number	11	10	10	10	9	9	9	8	9.5
Cluster number	22	20	20	24	14	18	15	22	19.375
Rule number	97	135	174	62	35	138	25	97	95.375
Misclassification ratio	23.50%	17.30%	18%	20.50%	9.10%	21.30%	11.90%	23.50%	18.14%
Averaged attribute differentiation ratio	18.18%	40%	13.33%	23.33%	29.63%	14.82%	29.63%	12.50%	22.68%
Un-Selected attributes		visittype[a]	orden	ptype	orden, visittype[a]	ptype,visittype	ptype,orden	ptype,orden,visittype	

[a] Better results.

Table 4
The tabulated results of experiments with different values of α

Threshold	MAX + MIN/2	MAX + MIN/3	MAX + MIN/4	MAX + MIN/5	MAX + MIN/6	MAX + MIN/7	MAX + MIN/8	MAX + MIN/9
α	2	3	4	5	6	7	8	9
MAX	0.31140577	0.3160365	0.2040913	0.37529672	0.22263813	0.2481348	0.3760075	0.1761336
MIN	4.87E-13	5.03E-09	1.13E-12	3.55E-10	3.19E-11	3.00E-09	4.30E-12	7.65E-10
Cluster number	24	16	22	23	24	25	22	20
Hierarchical cluster number	8	3	7	4	3	3	3	7
Rule number	6	54	29	9	36	12	35	42
Misclassification ratio	10.60%	11.40%	8.10%	7.30%	14.30%	7.10%	8.30%	5.60%
Averaged attribute differential ratio	30.30%	18.18%	24.24%	48.48%	21.21%	18.18%	21.21%	33.33%

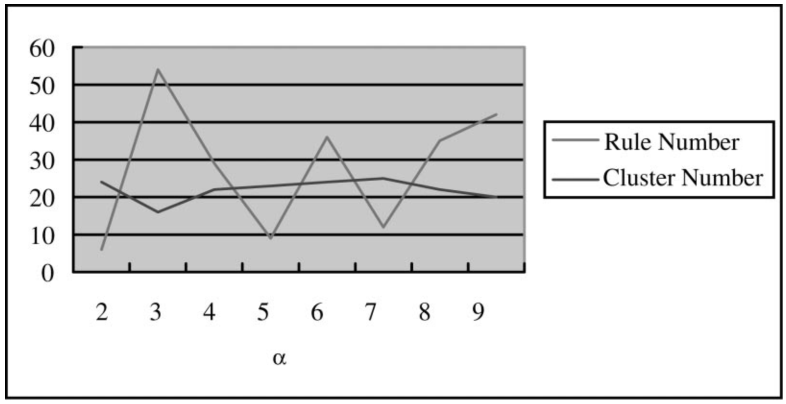


Fig. 15. The relationship between a and the cluster number in the experiments.

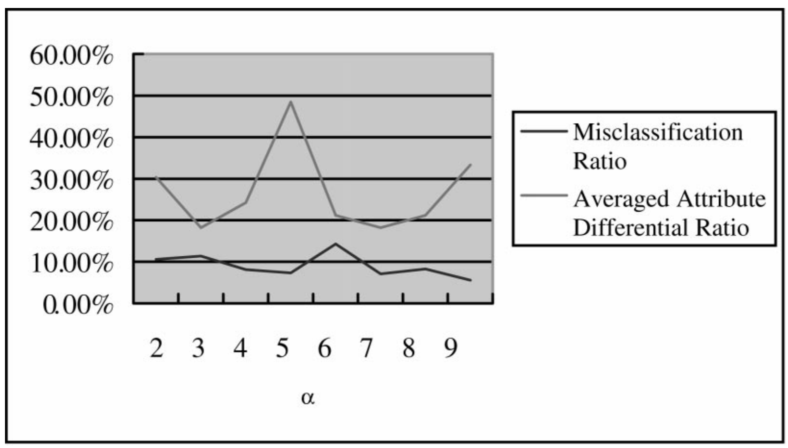


Fig. 16. The relationship betweenaand the averaged attribute differential ratios.

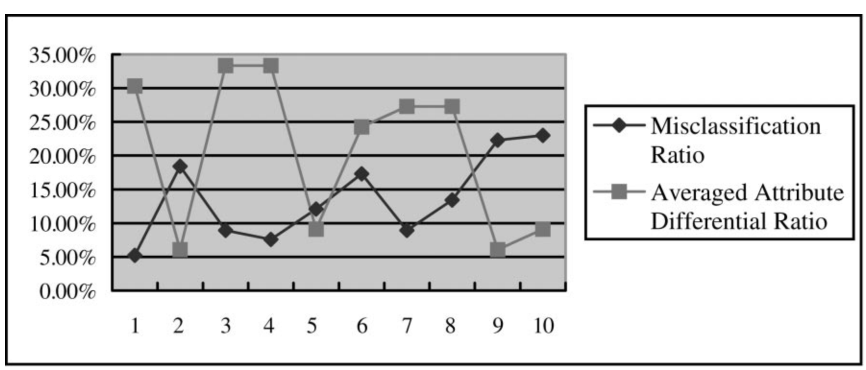


Fig. 17. The averaged attribute differential ratios for the 10 experiments conducted by a same human analyst.

Table 5
The tabulated results of 10 experiments conducted by a human analyst

	1	2	3	4	5	6	7	8	9	10
Cluster number	17	21	19	18	25	20	26	26	18	20
Rule number	68	41	149	13	47	47	31	18	110	163
Misclassification ratio	5.20%	18.40%	8.90%	7.60%	12.10%	17.30%	8.90%	13.40%	22.30%	23%
Averaged attribute differential ratio	30.30%	6.06%	33.33%	33.33%	9.10%	24.24%	27.27%	27.27%	6.06%	9.09%

The results are shown in Table 5 and Fig. 17. Based on this, we obtain the following observations: the averaged attribute differential ratios are similar when considering only those experiments with lower misclassification ratios. In other words, the approach preserves the stability of performance at the third place.

5. Conclusion

In this paper, we present a novel mixed-initiative synthesized learning approach for the better understanding of customers and the provision of clues for improving customer relationships based on different sources of Web customer data. This approach is a combination of H-LabelSOM, decision tree, cross-class analysis, and tacit experience of human analysts. The approach was implemented as a system called CRMiner and applied to the data of the Taiwan branch of the worldwide leading printer company. The analyzed results provide important clues for customer retention and the directions of customer-relationship improvement. We believe our approach can work for different enterprises as well in the future.

This approach is also evaluated for its stability in maintaining the quality of the results obtained when facing all possible different user-controlled settings, such as the factor of selected attributes/attribute number, the factor of cluster number, and the factor of cluster class assignment. The evaluation results are quite positive.

The future work of enhancing the approach is four-fold:

- Consider the importance of attributes to be represented by non-binary weights during analysis as indicated in Section 3.6.
- Devise more delicate ways of cross-class analysis.
- Build an expert system replacing the human efforts in reasoning out the clues for customer retention and directions of relationship improvement from the results of cross-class analysis.
- Use our approach to explore more aspects of CRM problems.

References

Customer Relationship Management, http://www.dci.com/crm

Saunders, C. & Meltzer, M. Driving Customer Retention, Development And Acquisition For Profit In The Insurance Business, http://www.crm-forum.com/crm_forum_white_papers/dcr/sld01.htm

Customer Retention Associates, http://www.customerloyalty.org/

Customer Retention Practices: Solutions, http://retention.harrisblackintl.-com/solutions/

Andreas, R. (1997). Alternative Ways for Cluster Visualization in Self-Organizing Maps, Proceeding of the Workshop on Self-Organizing Maps (WSOM97), Helsinki, Finland.

Andreas, R. (1999). LabelSOM: On the Labeling of Self-Organizing Maps, Proceedings of the International Joint Conference on Neural Networks (IJCNN'99), Washington, DC.

Andreas, R. (1999). Automatic Labeling of Self-Organizing Maps: Making a Treasure-Map Reveal its Secrets, Proceedings of the third Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'99), Bejing, China. LNCS / Lecture Notes in Artificial Intelligence, LNAI 1574, pp. 228–237, Springer Verlag.

UGAI97 Workshop, Building Classification Models: ID3 and C4.5, http://yoda.cis.temple.edu:8080/UGAIWWW/lectures/C45/

Data Mining Tools See5 and C5.0, http://www.rulequest.com/see5-info.html

Kohonen, T. (1997). *Self-organization maps*, (2nd ed.). Berlin, Heidelberg: Springer.

Back, B., Sere, K. & Vanharanta, H. (1997). Analyzing Financial Performance with Self-Organization Maps, Proceedings of Workshop on the Self-Organizing Map (WSOM'97), Espoo, Finland.

Vesanto, J. (1997). Data Mining Techniques Based on the Self-Organization Map, Thesis for the degree of Master of Science in Engineering.

Goser, K. F. (1997). Self organising maps for intelligent process control, Proceeding of the Workshop on Self-Organizing Maps (WSOM97), Helsinki, Finland.

Kaski, S., & Kohonen, T. (1996). Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In A.-P.N. Refenes, Y. Abu-Mostafa, J. Moody & A. Weigend, *Neural networks in financial engineering* (pp. 498–507). Singapore: World Scientific.