A Synthesized Learning Approach for Web-Based CRM

Wei-Lun Chang Soe-Tsyr Yuan
Business Intelligence Management Laboratory
Department of Information Management
Fu-Jen University, Taipei, Taiwan
E-mail: allen@im.fju.edu.tw yuans@tpts1.seed.net.tw

Abstract

The issue of customer relationship management emerges rapidly. Customers have become one of the important considerations to companies being built as well. Accordingly, customer retention is a very important topic. In this paper, we present a 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, CRMiner, is a combination of Self-Organization Maps, Hierarchical Automatic Labeling SOM, and Decision Tree, and Cross-Class Analysis. The objective of the approach is to segment data source 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.

1. Introduction

The relationship between companies and customers evolves into a critical research issue, and the concept "Customer Relationship Management (CRM)" was proposed frequently in the recent years. Moreover, CRM not only discusses traditional customer problems but provides an integrated solution to resolve internal problems in the enterprise, including "Customer Marketing", "Sale", "Customer Services", and integrate people, process and technology to form a revolutionary concept [1, 2]. 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" [3]. 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 issue [4].

On the other hand, although all enterprises have their own customers, they still have to discover potential target customer continuously. 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 of customer relationships.

For Web-based CRM, the sources of customer data are from the customer-Web interactions. In this paper, we present a novel 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 novel synthesized learning approach is a combination of Hierarchical Automatic Labeling SOM Clustering Method and Decision Tree.

This paper is organized as follows: Section 2 describes the architecture of the approach. Section 3 provides the descriptions of the three components of the approach, SOM, automatic-labeling SOM (LabelSOM), and decision tree. Section 4 demonstrates an example to show how the approach gains better understanding of customers and provides clues for improving customer relationships. Finally, a conclusion is made in Section 5.

2. Architecture

In this section, we provide the architecture of our synthesized learning approach and the properties of the approach. The architecture shown in Figure 1 shows there are fours major tasks associated with the synthesized learning approach:

1. Web-based Customer data is integrated for analysis. We used three sources of customer data (members, products purchased, member visit log) generated from the customer-Web interactions of the Taiwan branch of the worldwide leading printer company.
2. The centralized customer data is clustered by hierarchical automatic labeling SOM that generates visualization diagrams showing all clusters with labeled characteristic features automatically.
3. Human analysts assign classes (Normal, Defected, Possibly Defected) to clusters based on the labeled features of clusters.
4. The labeled features of clusters with class labels serve as a training sample for the decision tree, which learns the characteristics of Normal/Defected/Possibly Defected customers.

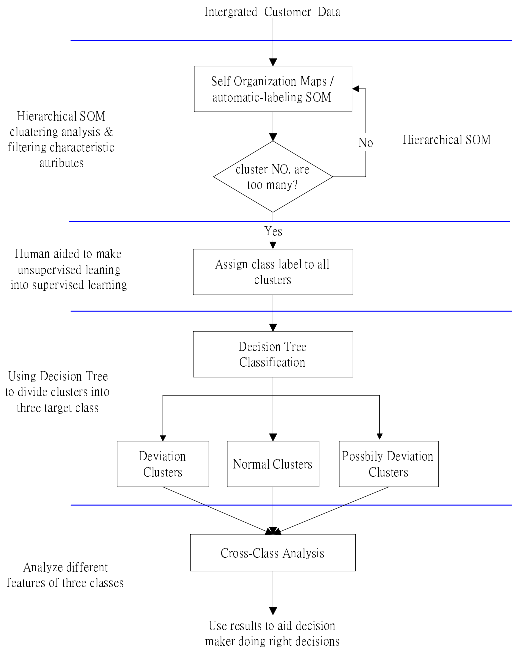

Figure 1： System Framework

The properties exhibited by the architecture are three-fold:

(1) Visualization clustering analysis method with automatic labeling

In general, businesses have huge amounts 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 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 previous statistical analysis approaches.

(2) Mixed-initiative:

This paper presumes that it is often the case 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.

(3) Customer Class characteristics are learned with the Decision Tree:

The labeled features of clusters with class labels enables the decision tree to learn the characteristics of Normal/Defected/Possibly Defected customers. Subsequent further comparison between cross-class characteristics can be made for providing clues to improve customer relationships

(3) Human-Computer Interaction

Human analysts play a vital role in the data mining process. It's the interactions that make data mining systems more humane and enable to obtain convincible and reliable results. In our approach, users have to set a threshold for controlling the size of clusters and make selections of attributes for subsequent analyses.

3. Methodology

The synthesized learning approach we present is a combination of three methods: visualization SOM [10], automatic-labeling [5, 6, 7], decision tree [8, 9]. In this section, the descriptions of these three methods are provided.

3.1 Visualization Self-Organization Maps

The visualization clustering method was proposed by Kohonen and become one of the most popular clustering methods recently. It is not only applied in resolving engineering problems but also employed to analyze data [11, 12, 13, 14]. Its strength is to project high dimensional data into two-dimensional grids.

The concept of visualization clustering method is close to the k-means clustering method. Assume each data record have many attributes called dimensions. The main ideas are two-fold:

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

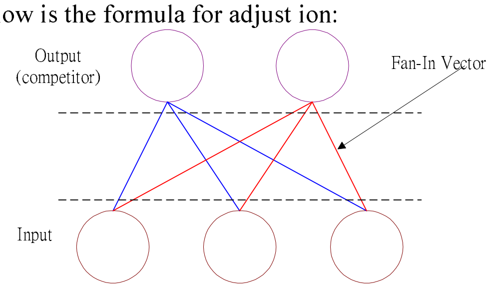

Figure2. Self-Organization Maps Concept

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \qquad \text{for each } i \in N_c(t),$$

t: time

Nc(t) : the neighborhood kernel function that identifies the set of neurons near to a best matching unit

α(t) : learning rate between 0 and 1

This formula aims to adjust all neighboring nodes which are near to the best matching unit. There exist two common functions for Nc(t), bubble function and gaussian function, and the most applied one is gaussian function. The reason why adjusting neighbor nodes as well is that it is necessary to make a whole cluster represented by neighboring neurons get closer to the data they contain in order to maintain the neighborhood relationship

.

3.2 Automatic Labeling Self-Organization Maps

Since T. Kohonen proposed the 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 can't automatic detect the boundaries of all clusters, and label the features of clusters generated. Rauber

Andreas [6] proposed a novel approach as LabelSOM which automatically label the features of clusters generated by SOM and ways of detecting the boundaries of clusters.

Automatic labeling aims to automatically filter the huge amounts of customer data records in a cluster into features. By labeling, we mean to label out the important attributes that matter the formation of the cluster. For example, after automatic labeling approach, only 5 attributes out of original 15 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. Therefore, we can quickly employ these features to analyze the characteristic of the cluster. Below is the formula of the distance measure, and Figure 3 demonstrates the concept of quantization error vector.

$$q_{i_k} = \sum_{x_j \in C_i} \sqrt{(m_{i_k} - x_{j_k})^2}, \qquad k = 1..n$$

qik : distance measure for the *kth* attribute in a given neuron *i*
Ci : the set of data records Xj projecting into a same neuron
Mik : the kth attribute value of the neuron model vector
Xjk : the kth attribute value of the Xj input vector

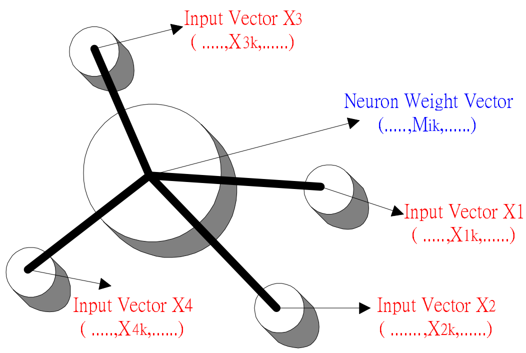

Figure3. The concept of the distance measure.

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.

3.3 Hierarchical Automatic Labeling SOM

In this paper, we apply the hierarchical concept to the LabelSOM approach and form the Hierarchical Automatic Labeling SOM method. The objective is to automatically reduce the size of the clusters constructed out of a set of data records until a threshold set by users is satisfied.

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, a problem should be considered：if irrelevant attributes are embedded in training examples, the accuracy of learned classifying rules is under question. Therefore, C5 [9] is the method we employ because it takes care of the problem of irrelevant attribute.

3.5 Cross-Class Analysis

The task of cross-class analysis enables the generation of valuable knowledge from data, to retain **Possibly-Defected** customers.

This task is comprised with two steps:

- As shown in Figure 4, with the intersection operation, the approach finds the intersected attributes/values within the classes Normal, Defect, Possibly Defected respectively, followed up the findings of intersected attributes/values between any two pair of classes.
- Comparing these 6 intersections, users are able to reason out the effective clues for improving the customer relationships. An example will be shown in Section 4.

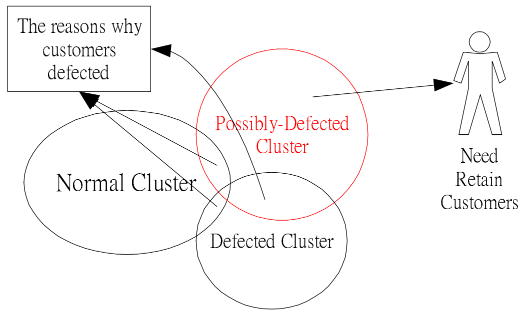

Figure 4. The Concept Of Cross-Class Analysis

3.6 Metrics

This section provide three metrics for monitoring the status of analysis:

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. CRMiner can reduce the size of clusters for the efficient generation of analyzed results.
2. **Attribute-Differential Ratio:**
   For any two classes, the attribute differential ratio is defined as the ratio of the number of different attributes in values between the two classes over the total number of attributes under analysis. Higher ratio implies it is much more worth of subsequent analysis for more information will be extracted. The formula is as follows:
   Attribute-Differential Ratio=D/T
   D: number of different attributes in values between the two classes over the total number
   T: total number of attributes
3. **Retention Rate:**
   The retention rate is computed right after cluster class assignment, and below is the formula:

Retention_Rate = (possibly-defected+defected) customer / total customers
If (Retention_Rate < default) then give-up providing suggestions

We will set the default value to 3% (the most common used in the companies) 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.

4. An Example and Discussion

After cleaning and pre-processing the data, we analyzed 3658 registered customer member data with 40 attributes from the Taiwan branch of the worldwide leading printer company. Each record includes a set of attributes integrated from three different sources of Web-based customer data represented by three different databases. The database "Member" recorded the basic information and profile of customer members. The database "Product" recorded members' purchasing history. The database "Visitlog" recorded members' behavior at the company's Web site, such as the first visit time, last visit time, if or not ordering e-news, and *etc*. We selected 11 useful attributes for analyzing. ("useful" means most important attributes valuable for analysis)

The results generated by our system, CRMiner (shown in Figure 5), have three parts: the result of the Hierarchical Automatic Labeling SOM unsupervised learning, the results of Decision Tree supervised learning, and the results of the cross-analysis over the characteristics of normal, defected and possibly defected clusters of customers for providing clues for gaining customer retention. Due to the limitation space, the last two parts of results are described below.

Figure 5. CRMiner System

Table1: The results of Decision Tree learning

<u>**Evaluation on training data (3658 cases)**</u>

Decision Tree		Rules	
Size	Errors	No	Errors
42	499(13.6%)	34	508(13.9%)

(a)	(b)	(c)	<-classified as
3087	*5*	*6*	(a): class 0
404	**34**	*2*	(b): class 1
89	*2*	**29**	(c): class 2

Table2. Cross-class intersected attribute/value

Class Name	Boundary	Attributes										
		memtype	sex	bdate	area	order	wtype	ptype	firstvist	lastvist	vistcnt	visittype
Normal (N)	Lower	0	0	0	0	0	0	0	0	0	0	0
	Upper	2	2	65	25	1	11	620	80	80	11	1
Possibly Defected (P-D)	Lower	0	0	---	0	0	0	0	0	0	0	0
	Upper	2	2	---	22	1	11	620	80	80	11	1
Defected (D)	Lower	0	0	0	0	0	0	0	0	0	---	0
	Upper	2	2	65	22	1	11	620	80	80	---	1
		Cross-class intersected attribute/value										
N/D	Lower	0	---	0	0	0	0	0	0	0	---	0
	Upper	2	---	65	22	1	11	620	80	80	---	1
N/P-D	Lower	0	---	---	0	0	0	0	0	0	0	0
	Upper	2	---	---	22	1	11	620	80	80	11	1
P-D/D	Lower	0	0	---	0	0	0	0	0	0	---	0
	Upper	2	2	---	22	1	11	620	80	80	---	1

Table 1 lists the results of Decision Tree learning. There are 3087 data in class 0 (Normal class), 34 data in class 1 (Possibly Defected class), and 29 data class 2 (Defected cluster). The error-rate of this decision tree learning is 13.6%. These results are learned from the 20 clusters of customer data constructed by the Hierarchical Automatic Labeling SOM. The average of attribute differential-ratio is 18.2% (each attribute differential-ratio for any two classes is 2/11=18.2%), which indicates it is worth of doing the further analysis, though the retention rate is 1.72%, which is lower than the default value 3%.

Table 2 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, members in D class live above south Taiwan, and the intersected values for the attribute "area" of N/D, N/P-D are both associated with members live above south Taiwan.
- Female members: there is nonempty intersected values for the attribute "sex" of P-D/D, and the averaged "sex" value over P-D/D members is nearly associated with female members.
- Members who use the Web site infrequently: members in the D class do not show any values for the attribute "visitcnt". This means members who do 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.

5. Conclusion

In this paper, we present a novel 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 novel synthesized learning approach is a combination of Hierarchical Automatic Labeling SOM Clustering Method, Decision Tree, and Cross-Class Analysis. The approach has been implemented into a system called CRMiner and was applied to the data of the Taiwan branch of the worldwide leading printer company. The analyzed results provide important clues for the directions of customer-relationship improvement. We believe our approach can work for different enterprises as well in the future.

Reference

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

[2] Chris Saunders and Michael Meltzer, Driving Customer Retention, Development And Acquisition For Profit In The Insurance Business, http://www.crm-forum.com/crm_forum_white_papers/dcr/sld01.htm

[3] Customer Retention Associates, http://www.customerloyalty.org/

[4] Customer Retention Practices : Solutions, http://retention.harrisblackintl.com/solutions/

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

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

[7] Rauber Andreas, 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, 1999. LNCS / Lecture Notes in Artificial Intelligence, LNAI 1574, pp. 228 - 237, Springer Verlag.

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

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

[10] Teuvo Kohonen, Self-Organization Maps, Springer, Berlin, Heidelberg, Second Edition, 1997.

[11] Barbro Back, Kaisa Sere, and Hannu Vanharanta, Analyzing Financial Performance with Self-Organization Maps, Proceedings of Workshop on the Self-Organizing Map (WSOM'97), Espoo, Finland, 1997.

[12] Juha Vesanto, Data Mining Techniques Based on the Self-Organization Map, Thesis for the degree of Master of Science in Engineering, 1997

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

[14] Samuel Kaski and Teuvo Kohonen, Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World, In Apostolos-Paul N. Refenes, Yaser Abu-Mostafa, John Moody, and Andreas Weigend, editors, *Neural Networks in Financial Engineering*, pages 498--507. World Scientific, Singapore, 1996.