Mobile Phone APP Software Browsing Behavior using Clustering Analysis

Chui-Yu Chiu, Ju-Min Liao and I-Ting Kuo
Department of Industrial Engineering and Management
National Taipei University of Technology
No.1, Sec. 3, Chung-Shiao E. Rd., Taipei, Taiwan 10643

Po-Chou Shih
Graduate Institute of Industry and Business Management
National Taipei University of Technology
No.1, Sec. 3, Chung-Shiao E. Rd., Taipei, Taiwan 10643

Abstract

With the development of information technology (IT), finding useful information existed in vast data has become an important issue. The most broadly discussed technique is data mining, which has been successfully applied to many fields. Data mining extracts implicit, previously unknown, and useful information from data. Clustering analysis is one of the most important and useful technologies in data mining methods. In this research, we utilize clustering analysis to analyze user loyalty of mobile phone APP software and use RFM analysis to describe the user behavior. By using the self-organizing map algorithm, the proposed system is expected to provide marketing survey industry with precise market segmentation for marketing strategy decision making and extended applications. According to the market segmentation, special offers to the varied consumer group can stimulate purchasing behavior and finally improve click-and-mortar conversion rate effectively.

Keywords
Data mining, Cluster analysis, RFM analysis, Self-organizing map algorithm

1. Introduction
As the information technology emerges in recent years, the internet changes not only business models but also the communication between enterprises and consumers. Internet is borderless, highly interactive, real-time and low cost. It integrates text, image, sound, video and animation. For enterprises, these characters create good channel for business activities such as advertising, marketing, bargaining and trade transactions.

Through the internet, small and medium-sized enterprises will have more competitiveness in the international market. At present, the internet has already been used widely. Companies provide quicker and better services using internets. The process of ordering, buying products and delivering services through the internet generates a lot of data that can be beneficial to business operations and management.

In recent years, the smartphones and the tablet computers are essential for mobile commerce. The contents and functions of mobile phone application software (APP) attract billions downloads through interne every day. Thousands of commercial APPs are implemented in the internet APP market. Huge data are generated when customers use the APP. A lot of interested information can be retrieved from the huge database using appropriate data mining technology which can be very beneficial in business aspects. The customer behavior is among the most valuable information.

In this research, data of APP browsing behavior are collected from an APP store in Taiwan and treated as customer behavior information. We utilize clustering analysis to analyze user loyalty of mobile phone APP software and use RFM analysis to describe the user behavior. The proposed system is expected to provide marketing survey industry with precise market segmentation for marketing strategy decision making and extended applications. According to the market segmentation, special offers to the varied consumer group can stimulate purchasing behavior and finally improve click-and-mortar conversion rate effectively.

2. Literature Review

2.1 Introduction of Data Mining

The science of extracting useful information from large data sets or databases is known as data mining. In the past, data mining has been referred to as knowledge management or knowledge engineering. Data mining tools help to automatically extract initial customer profile from data. With sufficient and proper database, data mining and business intelligence can provide new business opportunities.

Under acceptable computational efficiency limitations, data mining is a KDD process consisting of applying computational techniques that produces a particular enumeration of patterns or models over the data (Fayyad , et al 1996). Figure 1 illustrates an overview of the KDD process. Data mining plays an essential role in the knowledge discovery process. Detailed discussion of each step can be found in (Fayyad 1997).
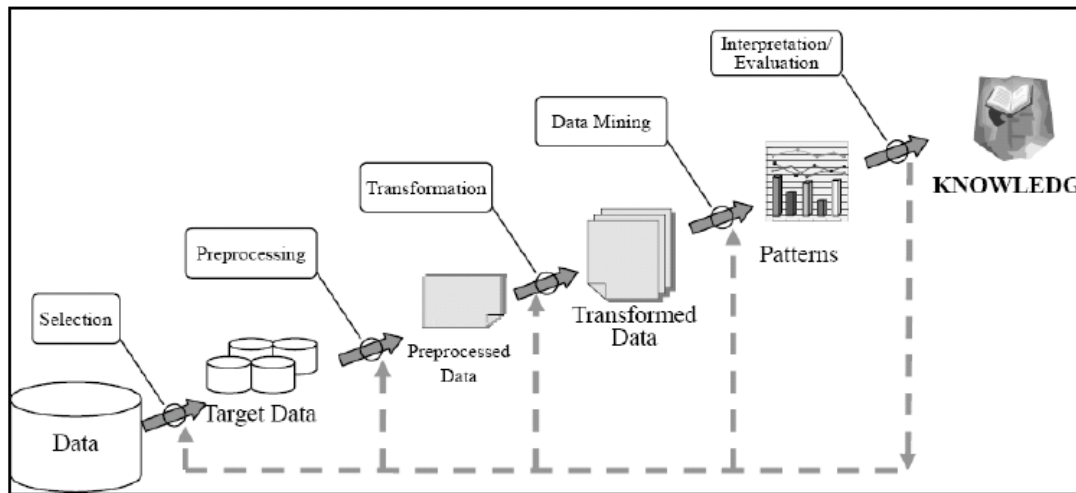

Figure 1. An overview of the steps that compose the KDD process (Fayyad, et al 1996)

2.2 Methods of Data Mining

The most known data mining methods includes (Han and Kamber 2001),(Mitra, Pal, and Mitra 2002),(Catherine Bounsaythip and Esa Rinta-Runsala):

1. Classification and Regression: Classification is a model function classifies a data item into one of several predefined categorical classes. There are several classification and regression techniques including decision trees, neural networks, Naïve-Bayes and nearest neighbor.

2. Clustering: This function maps a data item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models. Clustering techniques include k-means or k-nearest neighbours (k-NN), a special type of neural network called Kohonen net or Self-organizing map (SOM).

3. Rule generation: Association rule mining refers to discovering association relationships among different attributes. Dependency modeling corresponds to extracting significant dependencies among variables.

4. Summarization or condensation: This function provides a compact description for a subset of data. Data compression is significant because of the advantage it offers to compactly represent the data with a reduced database size, thereby increasing the database storage bandwidth.

5. Sequence analysis: It models sequential patterns, like time-series analysis, gene sequences, etc. The goal is to model the states of the process generating the sequence, or to extract and report deviation and trends over time.

3. Methodology

In this research, data are collected from an APP store in Taiwan and treated as customer behavior information for cluster analysis. The cluster analysis method is proposed as follows.

3.1 Research Framework

The RFM analysis framework is proposed to describe and analyze user browsing behavior. To segment the customer browsing behavior, we utilized SOM algorithm to perform data clustering. The research framework is shown in Figure 2.
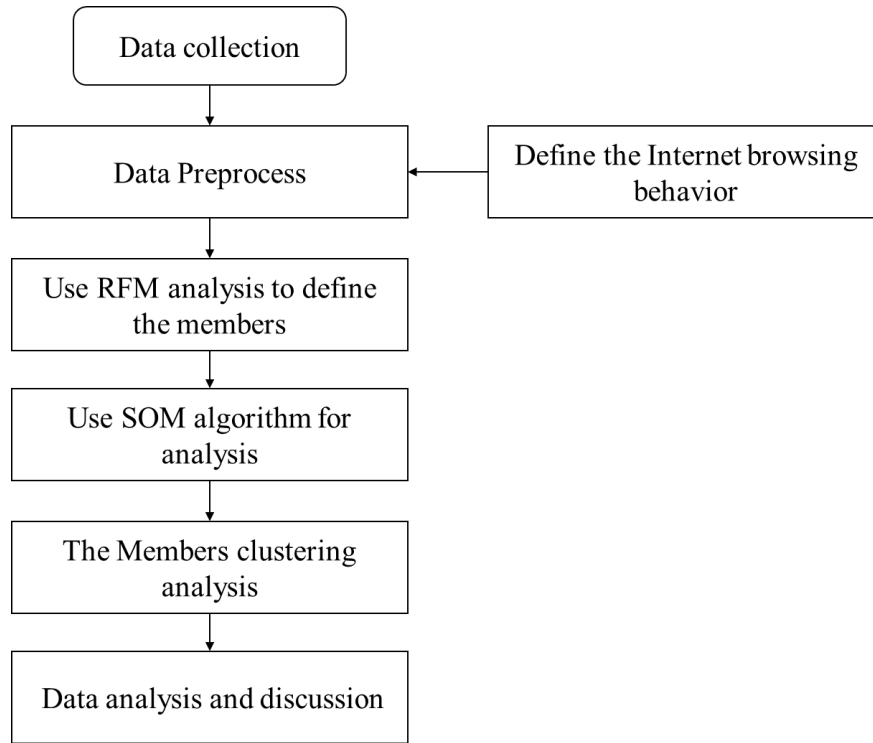
Figure 2 The research framework

3.2 Data Description

The most well-known data mining methods are Clustering, Classification and regression, Association rule discovery and sequential pattern discovery. According to Data mining techniques, we chose clustering technology to analyze customer market segmentation. Through market segmentation, industry can supply more suitable products or services for customer.

This study used Google Analytics to define users who liked to read gastronomic or traveling information on the APP. The log file of each customer represents his beginning and ending time of browsing and the contents browsed. When customers used the application software, Google Analytics record the member information through the post processor as shown in Table 1.

Table 1

ID	Article Hash	USER_UID	Time
1			
2			

1. ID: Serial number
2. Article Hash: Articles category
3. USER_UID: User ID
4. Time: Create time

3.3 RFM Analysis

Market segmentation divides a market into distinct subsets of customers, where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix. Behavior segmentation helps derive strategic

marketing initiatives by using variables that determine customer profitability (Ha and Park 1998), (Marcus 1998), (Saarenvirta 1998).

With customers' behavior data available, we need to decide which variables are to be used for analyzing. Among many behavior variables, RFM variables are well known to both researchers and practitioners. In order to completely understand customers, database marketers applied RFM variables to segment customer markets (Newell 1997). RFM is a simple method to provide a framework for understanding and quantifying customer behavior. In this research, we apply RFM analysis to identify customers browsing behavior. We define RFM characters as follows:

1. Recency (R) is defined as the time period since the last time browsed.
2. Frequency (F) is defined as the average number of browsing daily.
3. Monetary (M) is defined as the average time length of browsing during a day.

The comparison between traditional RFM and the proposed RFM model are shown in table 2.

Table 2. RFM models

Traditional RFM	Proposed RFM
R: Last purchase date (type: date) F: Count purchase M: Total money	R: Last browse day (type: day) F: Average number of browsing daily (type: number of times) M: Average time length of browsing during a day (type: seconds)

3.4 Self-organizing map algorithm (SOM)

SOM was first developed by Von der Malsburg (1973) and refined by T. Kohonen (1982). The SOM algorithm is in general a data compression technique that takes arbitrary high dimensional data as input and produces a lower dimensional map as output (Kohonen 1993). A self-organizing map consists of interconnected components called nodes or neurons. These nodes are connected to each other in a regular structure. There are many SOM structure alternatives (Kangas and Kohonen 1990).

A map has a two-dimensional lattice of neurons and each neuron represents a cluster. The adaptation process of SOM is an unsupervised competitive learning. All neurons compete for each input pattern. The winning neuron is selected from lattice. Only the winning neuron is activated. The winning neuron not only updates itself but also neighbor neurons to approximate the distribution of the patterns in the input dataset. After the adaptation process is complete, similar clusters will be close to each other.

The following describes the basic steps of SOM:

1. Randomly initialize all weights vector.
2. Select input vector $\mathbf{x} = [x_1, x_2, x_3, \ldots, x_n]$.
3. Compare $\mathbf{x}$ with weights $\mathbf{w}_j$ for each neuron j to determine winner.
4. Update winner and winner's neighbours.
5. Adjust parameters include learning rate and neighbourhood function.
6. Repeat from 2 until the map has converged or pre-defined number of training cycles have passed.

4. Case study

In this research, data of APP browsing behavior are collected from an APP store in Taiwan and treated as customer behavior information. These data were collected from February 2013 to July 2013. It totally includes 6253 data. We used Self-organizing map algorithm for the cluster analysis of the database. We use coefficient of concentration to resolve the choice of group number. The results are shown in table 3.

Table 3. Cluster coefficient of concentration

Groups	coefficient of concentration	Increment
10	202.0991	---
9	202.0991	0.0000
8	208.8954	6.7963*
7	233.3929	24.4975
6	272.6669	39.2740

According to the table, the changes between increments of from 8 groups to 7 groups is approximately 3.6 times. The Group 7 changed to the Group 6 is approximately increasing the changes between increments of from 8 groups to 7 groups is approximately 1.6 times. So we choose Group 8 as the best number of cluster. The clustering results using SOM is shown in table 4.

Table 4 Clustering Results of SOM.

Cluster centers	Customer Counts	R	F	M
C 1	1384	168.0601	0.0258	4.6672
C 2	691	201.7540	0.0215	3.8841
C 3	1033	183.7279	0.0224	4.0484
C 4	112	171.6295	0.2055	37.1892
C 5	970	78.1748	0.0338	6.1248
C 6	972	50.9586	0.0303	5.4755
C 7	928	22.4034	0.0434	7.8543
C 8	163	24.9166	0.3461	62.6382

Ha & Park (1998) proposed to position each customer cluster based on comparing average RFM values of each cluster with the total average RFM values of all clusters. If the center of a cluster is greater than the total average, then an upward arrow $\uparrow$ is assigned to that variable; otherwise, a downward arrow $\downarrow$ is assigned. According to this result, cluster characteristics could be analyzed and their strategic positions could be determined. The results would show in table 5.

Table 5. Categorized RFM value for each clusters

Cluster	Customer Counts	R	F	M	RFM Status
1	1384	168.0601	0.0258	4.6672	R↑F↓M↓
2	691	201.7540	0.0215	3.8841	R↑F↓M↓
3	1033	183.7279	0.0224	4.0484	R↑F↓M↓
4	112	171.6295	0.2055	37.1892	R↑F↑M↑
5	970	78.1748	0.0338	6.1248	R↓F↓M↓
6	972	50.9586	0.0303	5.4755	R↓F↓M↓
7	928	22.4034	0.0434	7.8543	R↓F↑M↑
8	163	24.9166	0.3461	62.6382	R↓F↑M↑
Total Avg.		116.8258	0.0419	7.5791	

Cluster 7 who has R$\downarrow$F$\uparrow$M$\uparrow$ and Cluster 8 who has R$\downarrow$F$\uparrow$M$\uparrow$ who can be considered as loyal ones who frequently use and make a large quantity of browse. Cluster 4 who has R$\uparrow$F$\uparrow$M$\uparrow$ who frequently use and make a large quantity of browse, but recently not to use. Cluster 5 who has R$\downarrow$F$\downarrow$M$\downarrow$ and Cluster 6 who has R$\downarrow$F$\downarrow$M$\downarrow$ are probably a new customer who recently use. Cluster 1, Cluster 2 and Cluster 3 who have R$\uparrow$F$\downarrow$M$\downarrow$ is likely to be vulnerable customers who have not use for a long time.

Among the eight clusters, Cluster 7 and Cluster 8 are selected as a target customer segment with the first priority, followed by Cluster 4. It is because that the effect to these target segments might become potentially greater than the effect to others from the RFM point of view.

In the context of data mining, clustering is a useful technique that partitions objects into clusters such that objects within a cluster have similar characteristics, while objects in different clusters are more distinct from another. Clustering is often an important initial step in the data mining process. Then, the performance of clustering is related to the level of the similarity within a group. Via RFM variables to indicate degrees of importance of customers. After customers had been clustered, target customers were identified through the strategic positioning of customer clusters. With the clustering information, it can help marketers to develop proper tactics for their customers. As a result, a competitive advantage could be generated from it.

5. Summary

In this research, data of APP browsing behavior are collected from an APP store in Taiwan and treated as customer behavior information. We utilize clustering analysis to analyze user loyalty of mobile phone APP software and use

RFM analysis to describe the user behavior. The proposed system is expected to provide marketing survey industry with precise market segmentation for marketing strategy decision making and extended applications. In the result of cluster analysis, each segment is identified as specific character in terms of customer loyalty. According to the market segmentation, special offers to the varied consumer group can stimulate purchasing behavior and finally improve click-and-mortar conversion rate effectively.

References

Fayyad, U., Piatetsky-Shapiro, G., and Smith, P., From Data Mining to Knowledge Discovery in Database, *American Association for Artificial Intelligence*, pp. 37-54, 1996.

Fayyad, U., Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases, *Scientific and Statistical Database Management*, in Proceedings of 9th International Conference, pp. 2-11, 1997.

Han, J., and Kamber, M., Data Mining: Concepts and Techniques. San Diego: Academic Press.

Mitra, S., Pal, S. K., and Mitra, P., Data Mining in soft computing framework: A survey, *IEEE Transactions on Neural Networks*, vol. 13, pp. 3-14, 2001.

Catherine Bounsaythip and Esa Rinta-Runsala, Overview of Data Mining For Customer Behavior Modeling, *VTT Information Technology*, printing

Ha, S. H., and Park, S. C., Application of data mining tools to hotel data mart on the intranet for database marketing, *Expert Systems with Application*, Vol. 15, pp. 1-31, 1998

Marcus, C., A practical yet meaningful approach to customer segmentation, *Journal of Consumer Marketing*, Vol. 15, pp. 494-504,1998.

Saarenvirta, G., Data mining to improve profitability, *CMA Management*, Vol. 72, pp. 8-12,1998.

Newell, F., The New Rule of Marketing: How to use one-to-one relationship marketing to be the leader in your industry, New York: McGraw-Hill Inc, 1997

Kohonen, T., Things you haven't heard about the self-organizing map, *Institute of Electrical and Electronics Engineers*, pp. 1147-1156 ,1993.

Kangas, A. J., and Kohonen, T., Variants of self-organizing maps, *Transactions on Neural Networks*, vol.1, no.1, pp. 93-99,1990.

Biography

Chui-Yu Chiu is an associate professor in the Department of Industrial Engineering and Management at National Taipei University of Technology, Taiwan. He received his Ph.D. degree in Department of Industrial and System Engineering from Auburn University. His research interests include data mining with applications and intelligent management systems

Ju-Min Liao received her Master of Science degree in the Department of Industrial Engineering and Management, National Taipei University of Technology from Taiwan. Her research interests include data mining, cluster analysis, RFM analysis

I-Ting Kuo received her PhD degree in the Department of Industrial Engineering and Management from, National Taipei University of Technology from Taiwan. Her research interests include data mining, cluster analysis, customer relationship management.

Po-Chou Shih received his PhD degree in the Graduate Institute of Industry and Business Management, National Taipei University of Technology from Taiwan. His research interests include data mining, cluster analysis, optimization process.