RESEARCH ARTICLE   **OPEN ACCESS**

Customer Clustering Based on Customer Purchasing Sequence Data

Yen-Chung Liu*, Yen-Liang Chen**

*(Department of Information Management, National Central University, Chung-Li, Taiwan 32001, R.O.C.)
** (Department of Information Management, National Central University, Chung-Li, Taiwan 32001, R.O.C.)

ABSTRACT
Customer clustering has become a priority for enterprises because of the importance of customer relationship management. Customer clustering can improve understanding of the composition and characteristics of customers, thereby enabling the creation of appropriate marketing strategies for each customer group. Previously, different customer clustering approaches have been proposed according to data type, namely customer profile data, customer value data, customer transaction data, and customer purchasing sequence data. This paper considers the customer clustering problem in the context of customer purchasing sequence data. However, two major aspects distinguish this paper from past research: (1) in our model, a customer sequence contains itemsets, which is a more realistic configuration than previous models, which assume a customer sequence would merely consist of items; and (2) in our model, a customer may belong to multiple clusters or no cluster, whereas in existing models a customer is limited to only one cluster. The second difference implies that each cluster discovered using our model represents a crucial type of customer behavior and that a customer can exhibit several types of behavior simultaneously. Finally, extensive experiments are conducted through a retail data set, and the results show that the clusters obtained by our model can provide more accurate descriptions of customer purchasing behaviors.
*Keywords***:** Customer clustering, Customer segmentation, Sequence data, NMF

I. INTRODUCTION

Customers are the most critical asset of companies, and customer churn and the subsequent loss of revenue must therefore be avoided. An increasing number of companies have invested resources in customer relations management (CRM) to quickly adapt their strategies and retain customers. CRM comprises four elements: customer identification, customer attraction, customer retention, and customer development[1]. The first task among them is customer identification, which can be further divided into two tasks according to purpose. The first task is target customer analysis, which involves finding customers who will be profitable for a specific company. The second task is customer segmentation, which involves clustering customers according to their similar characteristics.

The rapid development of e-commerce has led to more customer-related data, such as customer profile, transaction, and Web browsing data, being collected and stored in databases. By using data-mining techniques to analyze these data, we can discern the differences between customers and identify target customers. Clustering is the most commonly used data-mining method for customer identification.

Clustering is a common technique for data analysis in numerous fields including machine learning, pattern recognition, information retrieval, bioinformatics, and various areas of commerce. The task of clustering is to partition a set of objects into multiple groups, with the objects in the same group demonstrating high similarity and the objects in different groups display in glow similarity. This technique is usually applied to determine the major groups of customers in customer segmentation. For each customer group, appropriate sales plans or product development strategies can be suggested according to its characteristics.

In general, customer clustering is executed according to the different types of customer data involved, namely customer value data, customer profile data, customer transaction data, and customer purchasing sequence data. These data types are briefly introduced as follows.

Customer value data. In this approach, customers are clustered according to their value to the company. First, features are extracted from customer data, which can help us to determine to what degree a customer contributes to the company. A popular model used in this approach determines customer value according to three extracted features, namely recency, frequency, and monetary, and is thus referred to as the RFM model.

Customer profile data. The assumption behind this approach is that customers with the same background are more likely to display the same consumption behaviors than those with different backgrounds. Accordingly, features derived from demographic characteristics or background factors

are selected for clustering. For example, the clustering attributes could be gender, age, income, educational background, or the duration of Internet experience. After clustering, all customers are partitioned into groups with similar profiles or backgrounds.

Customer transaction data. The popularity of electronic transactions enables companies to easily obtain detailed transaction data for each customer, containing the items purchased and numbers of units. By aggregating these transaction data, we can learn the amount of each itema customer purchased. This aggregated data focuses on item content without considering temporal orderrelation among transactions, and thus represents a static description of consumer behaviors. The similarity between customers is decided by the similarity of purchases. The customers who purchased similar items should be assigned to the same group. In other words, the customers in each cluster demonstratesimilar static consumption behaviors. However, because the behavior ofa real customer constitutes a temporal sequence of transactions, the weakness of this approach is that it ignores the dynamic nature of customer behaviors. This defect may preventthe clusters obtained from being able to accurately describe the dynamic aspect of customer behaviors.

Customer purchasing sequence data. In this category, the behaviors of a customer are represented by a sequence of items purchased, where an item could be a real product or a higher-order conceptual item. Clustering by item sequence is based onsimilarity. Two customerswith similar item sequences should be assigned to the same group. The advantage of clustering on the basis of item sequence is that it canaccount for dynamic changes in sequence(i.e., the dynamic nature of customer behaviors is included for consideration). However, its weakness lies in that in reality a customer usually purchases more than one item in a transaction. The item sequences fail to represent the real purchasing behaviors of customers. Item sequence clustering is thereforeanunsuccessful approach to clustering customers on the basis of their purchasing behaviors.

In this research, our clustering will be conductedusing customer sequence data. However,in our model, a customer sequence is extended from a sequence of items to a sequence of itemsets. The reason forthis extension is because a customer may shop multiple times and buy a set of itemseach time. By accounting for this phenomenon, a sequence in our model can properly represent the purchasing sequence of a customer in the real world.

Additionally, unlike traditional research that assignseach customer to only one cluster, in our model we allow each customer to be clustered into multiple clusters or no cluster. The reason forthis difference is that in the traditional view,each cluster represents a group of customers with similar characteristics or behaviors, whereasin our model, each cluster represents a major type of purchasing behavior and each customer can simultaneously displayseveral types of behaviors. In other words, a traditional model focuses on showing whose consumption behaviors are similar, but our model can reveal typical types of customer behaviors and shows which of them aredemonstrated by an individual customer.

In this paper, our goal is touse customer sequence datato derivea set of clusters, which are sequences of itemsets. In addition, wealso show which clusters each customer participates in. The solution methodology consists of three phases. In the first phase, the original customer sequence data aretransformed into a matrix of patterns and customers, where a pattern is an elementary purchasing behavior. Then, in the second phase, we apply the nonnegative matrix factorization (NMF) method to transform the original matrix into two matrixes; the first matrix shows the content of each typical purchasing type and the second matrix shows the relationship between typical purchasing types and customers. Finally, the third phase assigns each customer into clusters according to the matrix of typical purchasing types and customers.

The remainder of this paper is organized as follows. In Section 2, the background and related works are reviewed. Our solution methodology is proposed in Section 3. Section 4 discusses how the experimentsare performed to validate the results. Finally, we providea conclusion in Section 5.

II. RELATED WORK

In Section 2.1, we review how previous research has applied clustering techniques to customer data. We then introduce the background of the NMF method in Section 2.2.

2.1.Customer clustering

As mentioned in Section 1, customer clustering is conducted according to different types of customer data, including customer value data, customer profile data, customer transaction data, and customer purchasing sequence data. In the following sections, we briefly introduce each type.

2.1.1 Customer value data

Most companies possess some VIPcustomers who contribute ahigher or more stable profitto the company than average customers do. Companies naturally want to identify those VIP customers and providethem with better care to avoid churn. Accordingly, a reliable methodof assessing customer value is essential. The RFM and customer lifetime value (LTV) models are most popularly used for customer value evaluation. Through these

Yen-Chung Liu. Int. Journal of Engineering Research and Application
www.ijera.com
ISSN : 2248-9622, Vol. 7, Issue 1, (Part -1) January 2017, pp.49-58

models, companies can learn which features should be extracted from the personal or transaction data of their customers and use them to assess customer value . It used three metrics, including the present value, the potential value, and customer loyalty, to evaluate customer value by using the LTV theory. A framework was designed to illustrate how these three metrics should be calculated. Finally, according to the values of these three metrics, they clustered customers into different segments[2],[3],Another study clustered credit card customers into different segments by using SOM technology based on previous repayment behavior and RFM behavioral scoring predicators[4],Other researcherscompared the advantages and disadvantages of the customer clustering results derived withthree approaches, RFM ,chi square automatic interaction detection , and Logistic regression, for two different client groups, members of an e-mail-order retailerand a nonprofit organization. This approachis usually applied to target customer analysis, with the goalof finding customers who possess certain characteristics[5].

2.1.2 Customer profile data

In this line of research, the features used are usually collected from data concerning the backgrounds or demographic characteristics of customers. One researcher clustered sightseeing visitors to Cape Town based on features selected from tour characteristics, opinions about Cape Town, and demographic characteristics[6], another researcherevaluated 40 potential customers of a drink company on the basis oftheir degree of socialization, leisure, achievement and other indicators, and appliedthree different data mining methods to compare the clustering results, including a support vector machine, self-organizing feature map (SOFM), and k-means method[7], the other researcher proposed a two-stage customer clustering methodology combined with SOFM and k-means based on features including user experience, services provided in an electronic goods store, and demographic characteristics[8], and other researcher proposed an integrated method of profitable customer segmentation for a motor company according to features collected from demographic, customer satisfaction, and accounting data[9],The other researcher developed a soft clustering method that uses a latent mixed-class membership clustering approach to classify online customers according to four types of features, namely consumer network usage, satisfaction with service, buying behaviors, and demographic characteristics; notably, because the clustering resultsare soft rather than hard, a customer may belong to multiple groups[10].

This line of study is helpful for understanding into what groups a customer base can be divided, thereby providing support for product planning and market promotion. However, because the characteristics used for clusters may be changed depending on the application, this approach should be custom designed according to the specific demands of the application.

2.1.3 Customer transaction data

Earlier studies have performed customer segmentation on the basis of demographic characteristics or background information. However, clustering customers according to profile data is most likely no longer useful in the current, increasingly complicated world. In recent years, the use of customer transaction or behavioral data for customer clustering has become a popularresearch subject. Trading can be conveniently performed through the Internet, a process that enablesthe easy recording of customer transactions. Because these transaction data can reveal customer preferences and needs, research on the use of transaction data in customer segmentation is gradually becoming a trend.

some researchers developed a market segmentation methodology based on product-specific variables such as items purchased and the associated monetary expenses from the transactional histories of customers[11],other researchers represented each transaction as a vector, with each vector element corresponding to an item and 0 and 1 representing the absence and presence of the item in a transaction, respectively ; they then converted all vectors into an association matrix, in which they found similarities between clients and obtained clustering results[12]. The other researchers represented the purchasing behaviors of each customer through a vector of items and determined the distance between two items through the product hierarchy; after the distance between two customers was computed by determining the distance between the corresponding vectors, customers were clustered according to these distances[13],the other researchers represented customer data by using a vector of items with quantity values, determined the similarity of two customers by the amount of items, and clustered customers according to their similarities[14].

These studies have determined clusters on the basis of similarities between customers, whom were decided according to the similar characteristics and differences in quantity of the items they purchased. They ignored the fact that the purchasing behaviors of a customer are constituted of a sequence of purchasing decisions made over time. Numerous patterns can be identified from the entire purchasing history of a customer, information crucial for describing and discriminating customer behaviors. If this dynamic aspect is not considered, the clusters obtained may be unreliable in practical situations.

2.1.4 Customer sequence data

The advantage of this approach is that the dynamic aspect of customer behaviors is included by formattingthe data as a sequence of items purchased over time. Accordingly, the clusters obtained inthis model are more reasonable than those obtained usingprevious static models. Several studies haveused this model[15],[16].In general, item sequences may be appliedin areas as varied as genomic DNA, protein sequences, Web browsing records, and transaction records.

Some researcher proposed a new measure to compute the similarity between two sequences and developed a hierarchical clustering algorithm to cluster sequence datasets such as protein sequences, retail transactions, and web-logs[15],other researcher proposeda method combining model-based and heuristic clustering for categorical sequences. In the first step, the categorical sequences are transformed throughan extension of the hidden Markov model into a probabilistic space, where a symmetric Kullback–Leibler distance can operate. In the second step, the sequences are clustered using hierarchical clustering on the matrix of distances[16].

In the study of customer sequence data clustering, an item could be a real, specific product at the lowestlevel or a conceptual item at a higher level. Although item sequence clustering seems applicable to customer purchasing data, the behaviors of a customer areconstituted by a sequence of itemsets rather a sequence of items. In a real user transaction, a user may buy several items at a time;it is unrealistic to assume that only a single item is purchased in each transaction. For this reason, the item sequence clustering approach cannot be applied to real customer purchasing data.

Traditional research assigns each customer to only one cluster: all customers are partitioned into a complete set of nonoverlapping groups, where each group possesses the same behaviors. By contrast, in our model, we derive a certain set of clusters from the behavioral data of all customers. Each cluster denotes a major type of purchasing behavioror lifestyle. We accordinglyallow each customer to be included in multiple clusters or no cluster, because people can possess diverse lifestyles. For example, a student with a photography hobby may possess both the lifestyle of a student and that of a photographer. In other words, a traditional model focuses on partitioning customers into groups with similar behaviors, but our model focuses on revealing typical types of customer behaviors.

2.2 NMF

Each element in the shopping sequence usually contains more than one item; therefore, a traditional sequence clustering method is not applicable to customer purchasing sequence clustering. To remedy this problem, the original customer sequence data aretransformed into a matrix of elementary patterns and customers, where anelementary pattern corresponds to an elementary purchasing behavior. Essentially, an elementary pattern is a binary tuple, such as (a, b), which means item b will be boughtdirectly after the purchase of item a. We then apply the NMF method to transform this matrix into two matrixes; the first matrix shows the content of each typical purchasing type and the second matrix shows the relationship between typical purchasing types and customers.Specifically, the first matrix shows how each typical purchasing type is formed by elementary patterns, whereas the second matrix shows how a customer consumption behavior is constituted by typical purchasing types.The NMF method can help us to extract typical purchasing types from customer purchasing sequence data.

NMF is a method of multivariate analysis and linear algebra that is commonly used in latent semantic analysis. A nonnegative matrix V can be decomposed into two matrixes, W and H, with $V = WH$, where all the values of the elements in the three matrixes are not negative. Here, matrix V is of the size $m \times n$, matrix W is of the size $m \times r$, andmatrix H is of size $r \times n$. The value of r is set by the user, and is typically r«m and n. A small value of r can greatly reduce the matrix sizes, which facilitates further analysis.

Some researcherused the NMF method for document clustering. They first found all keywords from documents and then built a matrix V of these keywords and documents. By using NMF, they decomposed matrix V into W and H. In their study, W wasthe matrix of keywords and themes. Similarly, H wasa matrix of themes and documents, which demonstrateshow a document is formed by themes. They assigned a document to the theme with the largest value in the row of this particular document in H. Theyalso indicatedthat NMF was more suitable forfindingthe semantic characteristics of themes than other methods such as singular value decomposition, because NMF did not require that the themes be orthogonal toeach other. Without this restriction, NMF can focus on finding the themes with the best clustering effect [17].

Other researcherapplied the NMF method todocument summarization. They first built a matrix V of keywords and sentences, and then decomposed it into two matrixes W and H, where H was a matrix of topics and sentences. Finally, they chose the sentences with the highest score for each topic as the summary of the document.[18]. Because customers may have diverse lifestyles, clusters canoverlap with each other and a customer may belong to multiple clusters. NMF is suitable for solving our situation because it can indicatemultiple typical purchasing

types and assign a customer to several types simultaneously.

III. METHODOLOGY

The purchasing sequence of a customer reveals their preferences or behaviors. When two customers have different preferences, the characteristics of their purchasing sequences will likelydiffer. The types and quantities of items a customer buys arefrequently used to represent the characteristics of theirpurchasing behaviors. In addition tothis static nature of purchasing behaviors, we can consider how to draw the dynamic features of customers to describe their behaviors. How theitems purchased change over time is a key dynamic aspect that can be discovered from purchasing sequence data. If this dynamic feature is also used to discriminate customer behaviors, we can expect a more reliable clustering result from a more complete consideration. Accordingly, we analyze what patterns are frequently repeated in customer purchasing sequences. These repeated patterns arecrucial for discerning different user behaviors. When users have different preferences, different repeated patterns occur in their respective purchasing sequences. In other words, when two sequences differ, they must have different repeated patterns, and vice versa. We can thus use the repeated patterns in sequences to discriminate customer behaviors.

In this paper, we adopt elementary patterns to describe and discriminate customer behaviors.An elementary pattern is a binary tuple, such as(a, b), which means that item b will be bought directly after the purchase of item a. We use elementary patterns to describe customer behaviors for several reasons. First, they can be easily collected. Second, becausetheir form is simple, their usecan reduce the number of combinations that we must consider, which in turn reduces the computation cost. Third, they contain enough information to represent user behaviors, including the information about what items we buy and in what quantity and order.

Our algorithm contains three phases. First,

the original customer sequence data aretransformed into a matrix of elementary patterns and customers, where an elementary pattern is a repeated unit purchasing behavior in customer transaction data. The second phase applies the NMF method to transform the above matrix into two matrixes, where the first matrix shows the content of each typical purchasing type and the second matrix shows the relationship between typical purchasing types and customers.The third and final phase assigns each customer to clusters according to the matrix of typical purchasing types and customers. In the following subsections, we describe each phase in detail.

3.1 Data transformation

The goal of this phase is to build a matrix of elementary patterns and customers from customer purchasing sequence data. We first providethe following definitions.

Definition 1. Let $I = \{i_1, i_2, \dots, i_r\}$ denote the set of all items occurring in customer sequence data. The j-th transaction of customer c_m can then be denoted as $T_m^j = \{i_1^j, i_2^j, \dots, i_s^j\}$, $1 \leq s \leq r$. $T_m^j \subseteq I$, $|T_m^j| = s$, $|T_m^j| \geq 1$.

Definition 2. Let T_m^j and T_m^{j+1} be two consecutive transactions of customer c_m in sequence, and let i_x^j denote an item in T_m^j, and i_y^{j+1} denote an item in T_m^{j+1}. Then, $\{i_x^j, i_y^{j+1}\}$ forms anelementary pattern for $1 \leq x < |T_m^j|$ and $1 \leq y < |T_m^{j+1}|$. Let P_m^{j+1} denote the set of all elementary patterns generated fromT_m^j and T_m^{j+1}. Then,$|P_m^{j+1}| \leq |T_m^j| \times |T_m^{j+1}|$.

Definition 3. If customer c_m has n transaction in his purchasing sequence, all the elementary patterns of customer c_m can be denoted by P_m, where $= (P_m^2 \cup P_m^3 \dots \cup P_m^n)$.

Example. Assume that customer c_m has five transactions in sequence as follows. A total of 16 elementary patternsare generated as a result.

Table 1. Generating elementary patterns throughtransaction data

The items purchased at the j-th transaction	Elementary patterns generated
$T_m^1 = \{i_a, i_c, i_e\}$	
$T_m^2 = \{i_b, i_c\}$	$P_m^2 = \{(i_a, i_b), (i_a, i_c), (i_c, i_b), (i_c, i_c), (i_e, i_b), (i_e, i_c)\}$
$T_m^3 = \{i_c, i_d\}$	$P_m^3 = \{(i_b, i_c), (i_b, i_d), (i_c, i_c), (i_c, i_d)\}$
$T_m^4 = \{i_h\}$	$P_m^4 = \{(i_c, i_h), (i_d, i_h)\}$
$T_m^5 = \{i_b, i_c, i_e, i_f\}$	$P_m^5 = \{(i_h, i_b), (i_h, i_c)), (i_h, i_e)), (i_h, i_f)\}$

However, because (i_c, i_c) occurs twice, a total of 15 distinct elementary patterns are generated.In other words, we have $P_M = P_m^2 \cup P_m^3 \cup P_m^4 \cup P_m^5 =$
$\{(i_a, i_b), (i_a, i_c), (i_b, i_c), (i_b, i_d), (i_c, i_b), (i_c, i_c), (i_c, i_d), (i_c, i_h), (i_d, i_h), (i_e, i_b), (i_e, i_c), (i_h, i_b), (i_h, i_c), (i_h, i_e), (i_h, i_f)\}$.
Definition 4. Let the set of customers be denoted as $C = \{c_1, c_2, \dots c_p\}$. The elementary patterns of all customers

Yen-Chung Liu. Int. Journal of Engineering Research and Application
ISSN : 2248-9622, Vol. 7, Issue 1, (Part -1) January 2017, pp.49-58

www.ijera.com

can then be denoted by P_{All}, where $P_{All} = \{P_1 \cup P_2 \ldots \cup P_p\}$.

Definition 5. Let PM denote the matrix of elementary patterns and customers. Specifically, $PM_{x,y}$ is the number of times that pattern x appears in the sequence data of customer c_y, where $1 \leq x \leq |P_{All}| \cdot 1 \leq y \leq p$.

Table 2. Matrix of elementary patterns between customers

Pattern/Customer	C_1	C_2	...	C_{p-1}	C_p
P_1					
P_2				3	
...					
P_{q-1}				2	
P_q					

P_{q-1} appears twice in C_{p-1} transaction record P_2 appears 3 times in C_{p-1} transaction record

Definition 6. Let PM' be the normalized matrix of PM. Specifically, $PM'[x,y] = \frac{PM[x,y]}{\sum_{i=1}^{|P_{All}|} PM[x_i,y]}, (1 \leq x \leq |P_{All}|, 1 \leq y \leq p)$.

We must normalize the matrix because customers who purchase large amounts of items tend to display high frequencies in all elementary patterns; the same is true for customers who buy comparatively less. By normalizing these values with respect to the total frequency of a customer, we can reveal the relative frequencies of patterns of this individual customer, which represent their true preferences.

Table 3. Normalized matrix of PM

Pattern/Customer	C_1	C_2	...	C_{p-1}	C_p
P_1					
P_2				0.6	
...					
P_{q-1}				0.4	
P_q					

The proportion of P_{q-1} in C_{p-1} transaction record is 0.4

The proportion of P_2 in C_{p-1} transaction record is 0.6

We want to assign customers $c_1, c_2, \ldots c_p$ to clusters $G_1, G_2, \ldots G_q$ through analyzing matrix PM'. If the behavior of customer c_m is similar to several clusters, such as G_1, G_3, G_4, we can make $c_m \in G_1, G_3, G_4$.

3.2 Matrix decomposition through NMF

In the first phase, we retrieved the elementary patterns from customer purchasing sequence data, hoping to assess the similarity of the customers through determining the differences between the elementary patterns. Because the

elementary patterns are numerous and interdependent, they cannot feasibly be used todirectly evaluate the differences between customers. Instead, we apply the NMF method to transform the original matrix $PM' = $ V into two matrixes, where the first matrix W shows the content of each typical purchasing type and the second matrix H shows the relationship between typical purchasing types and customers.

After applying NMF to the matrix PM' (V), we obtaintwo matrixes, W and H, by the relation that

$PM_{|P_{all}|\times p} = W_{|P_{all}|\times r} \times H_{r \times p}$. Here, r is the number of typical purchasing types specified by a user. Usually, we set r as equal to the number of groups asked by user. Matrix W enables us to understand how each typical purchasing type (group characteristic[GC]) is formed by elementary patterns. Similarly, matrix H shows us how the consumption behaviors of each customer areconstituted by typical purchasing type (GC). In the following figure, we set r=2. Matrix V can then be transformed into two matrixes.
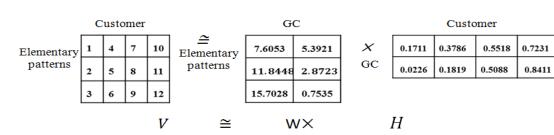

Fig. 1. Relationships amongmatrixes V, W, and H

From matrix H, we learn how the four customers are related to two different group characteristics (typical purchasing types).

Additionally, matrix W tells us how each group characteristic is constituted by elementary patterns.

3.3 Clustering

Table 4. Example of how to generate a clustering threshold

Group \Customer	...	C_{p-3}	C_{p-2}	C_{p-1}	C_p	Customer's GC value >0	Sum of GC value	Average
G_1	0	0	0	0.00136	0.02287	3873	29.15666	0.00752
G_2	0.00494	0.00755	0.00456	0.01724	0.01857	5652	59.5683	0.01053
...	0.00011	0.00070	0	0.00047	0	1651	3.96295	0.00240
G_{r-1}	0.00012	0.00001	0.00009	0	0	2611	5.34028	0.00204
G_r	0.00107	0	0.00258	0.00603	0.01154	3984	21.50367	0.00539

Matrix H shown in Fig. 1, shows us thatthe consumption behaviors of customer $C_p$areformed by weight 0.02287 on G_1 group characteristics, weight 0.01857 on G_2 group characteristics, weight 0.01154 on G_r group characteristics, and zeros on others. How can we assign customer C_p to groups? Intuitively, if a customer has a comparatively larger weight in a certain group, we should assign the customer to this particular group. Therefore, among all customers with positive weights in a group, we assign the customers whose weights are in the p-percentile into this cluster, where p is set according to the user.

On the basis of this clustering strategy, a customer is clustered into multiple groups or no group. Notably, we use the p-percentile as the threshold; thus, every group is nonempty because there must besome customers with weights in the p-percentile.

IV. EXPERIMENTS

To evaluate the performance of the algorithm, we mustdefine a metric to measure the

qualityof aclustering result. A group G_i of customers containsa total ofu_i customers and n_i elementary patterns. Let $p_{i1} \sim p_{iN}$ denote the most frequent N elementary patterns in group G_i. Our hypothesisis that fora group witha more favorableclustering effect, more customers shoulddisplaysimilar behaviors. In other words, there should be more customers in the group possessing top N frequent patterns. We use

$$PPP_i = \sum_{j=1}^{N} \frac{\text{the number of customers in } G_i \text{ possessing } p_{ij}}{u_i}$$

to represent the average proportion of customers in group G_i,possessing patterns $p_{i1} \sim p_{iN}$. Additionally, to measure the interdistance between groups, we define

$$PDP_i = \frac{\sum_{j=1}^{r}(1 - \frac{\text{the number of TOP N pattern overlap between } G_i \text{and } G_j}{N})}{r-1}, \text{ and } j \neq i$$

to measure the difference between group G_i and other groups. For example, assume that the number of clusters is four, and among the top three patterns, there are twocommon patterns between G_1 and G_2, onecommon pattern between G_1 and G_3, and two

common patterns between G_1 and G_4. We therefore derive $PDP_1 = \frac{[(1-\frac{2}{3})+(1-\frac{1}{3})+(1-\frac{2}{3})]}{4-1} = 0.4444$. An increasing value of PDP indicates fewer common patterns between different groups. In other words, intergroup differences become larger.

To integrate these two indexes, we use the following index to measure the clustering effect of group G_i: $CI_i = PPP_i \times PDP_i$. A larger index is considered more favorable. Finally, we use $CI = \frac{\sum_{i=1}^{q} CI_i \times u_i}{u_1 + u_2 + \cdots + u_q}$ to measure the total clustering effect, which is the weighted average of CI_i in all groups. When the value of CI is large, the purchasing patterns of the customers in the same group are more similar with one another and the purchasing patterns of the customers in different groups are more dissimilar with one another. When the value is 1, all customers in each group display the same top N patterns, and the top N patterns of all groups are different.

The data set used in the experiment is public customer purchasing sequence data collected from a store on 2001/12/27 and 2002/12/31. The data set contained information for 7,824 customers [19]. Among them, 2,096 customers were one-time buyers; thus, we remove them from consideration. A total of 5,728 customers and 54,537 transactions are eventually included in the experiment. The items used in the experiment are higher-level conceptual items, to avoid an exceptionally high number of patterns with low support. As a result, we obtain 110 items and generate 10,401 elementary patterns. We use the software MATLAB toolbox to perform NMF algorithm. This toolbox was released by *Source Code for Biology and Medicine* [20].

The contrast algorithm used in the experiment is the k-means algorithm, in which each customer is viewed as a data point. Because we have 10,441 elementary patterns, each data point has 10,441 dimensions. However, when the number of dimensions is excessively large, a dimensionality problem will occur and render the running time intractable. We therefore choose Q most frequent elementary patterns in place of the original 10,441 dimensions. To find the most appropriate value of Q, we test different potential values, including 5, 10, 15, 20, 30, 35, 40, 50, 100, and 200. The results indicate that when the value of Q is set to 30 or 40, we have the optimal clustering effect. In this experiment, we therefore decide to set Q to 30.

We additionally set k as 5, 10, 15, 20, 25, and 30, which are the numbers of clusters used in the experiment. We use a k-mean algorithm to cluster the 5,728 data points with 30 dimensions into k clusters. The distance between data points is computed by using a formula involving the square root of each

element of the array. We use the Python programming language to perform a k-means algorithm.

In the experiments, we set the thresholds of NMF to 50, 60, and 70. In other words, among all customers with positive weights in a group, we assign the customers whose weights are in the 50th, 60th, and 70th percentiles to this cluster. In addition, we set the number of groups as 5, 10, 15, 20, 25, and 30 for NMF and k-means. Furthermore, for the top N frequent patterns, we set N as 3 and 6. The two algorithms are executed 30 times for each combination of parameters. The results for when N =3 and N = 6 are presented in Tables 5 and 6, respectively.

Table 5. CI of TOP 3 patterns through k-means and the NMF method

K	K Means	NMF thr=70	NMF thr=60	NMF thr=50
5	0.228	0.225	0.238	0.256
10	0.282	0.283	0.297	0.310
15	0.314	0.310	0.327	0.336
20	0.329	0.328	0.344	0.355
25	0.340	0.340	0.357	0.366
30	0.353	0.350	0.364	0.373

The results in Table 5 indicate that when N=3, NMF has similar performance as that of k-means if the threshold is 70. However, if the threshold is 60 or 50, NMF has a more favorable performance than k-means does. In addition, the results show that when the threshold is lower, the clustering effect of NMF improves, because when the threshold is lower, we use a higher standard to assign a customer to a group. This causes the customers in the same group to exhibit more similar behaviors.

The results in Table 6 indicate that when N=6, NMF has a more favorable performance than k-means in all different thresholds. In addition, the results show that when the threshold is lower, the clustering effect of NMF improves.

Table 6. CI of TOP 6 patterns through k-means and the NMF method

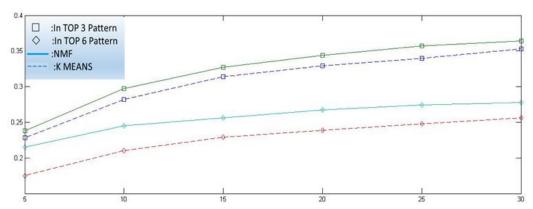| K	K Means	NMF thr=70	NMF thr=60	NMF thr=50
5	0.175	0.193	0.215	0.229
10	0.21	0.228	0.245	0.256
15	0.229	0.244	0.256	0.268
20	0.239	0.254	0.267	0.277
25	0.248	0.263	0.274	0.284
30	0.256	0.268	0.278	0.288

Fig. 2. Change in CI from k=5 to 30 in k-means and NMF with thr=60%

The results displayed in Fig. 2 compare the performance of k-means and NMF with the threshold of 60. In this comparison, N is set to 3 and 6 and the number of groups is set to 10, 15, 20, 25, and 30. The results indicate that when N is smaller, the clustering performance is more favorable, because when N is small, only the few most frequent patterns are used for computing group cohesion.

V. CONCLUSION

In the past, the behaviors of a customer were usually represented by a sequence of items purchased. Clustering by item sequence is instead based on the similarity of item sequences. The advantage of clustering based on item sequence is that the dynamic nature of customer behaviors is included for consideration. However, its weakness is that, in reality, a customer usually purchases more than one item in a transaction. The item sequences fail to represent the real purchasing behaviors of customers. Therefore, this research studied the customer clustering problem on the basis of transaction sequence rather than item sequence.

Unlike traditional research that assigns each customer to only one cluster, our model assigns each customer to multiple clusters or no cluster. The reason for this difference is that in the traditional view, each cluster represents a group of customers with similar characteristics; by contrast, in our model, each cluster represents a typical type of purchasing behavior and each customer can manifest several types of behavior simultaneously. We thus apply a different model, NMF, to cluster the transaction sequence data of all customers. The results obtained include two matrixes, where the first matrix shows the typical types of consumption behaviors and the second matrix shows which of these types a customer exhibits.

In the experiments, we compare our model with a traditional k-means algorithm. The results indicate that the clusters obtained by our model have higher CI values than do those obtained by k-means. In other words, compared with the clusters obtained by k-means, in our model the purchasing patterns of the customers in the same group are more similar and the purchasing patterns of the customers in different groups are more dissimilar.

In the future, this research can be extended in the following areas. First, in this work, the NMF model is used for clustering customer transaction sequence data. However, NMF is a general model for matrix decomposition and it may not be the most fitting for the intended work. In the future, we may attempt to custom design other models that are more appropriate for this situation. Second, in this work, we extract elementary patterns to represent the original transaction sequence data. Because elementary patterns are only one type of information that can be extracted from sequence data, we may attempt other methods to extract important information from the original sequence. We believe there is a chance that other extracted information can provide the same or even an increased ability to build the model. Finally, typical types of customer behaviors are first introduced in our clustering work. The problem of how to find these types is more interesting than the clustering itself. In the future, we will propose a study that focuses entirely on finding typical types of customer behaviors in customer transaction sequence data.

REFERENCES

[1]. Ngai, E. W., Xiu, L., & Chau, D. C, Application of data mining techniques in customer relationship management: A literature review and classification, *Expert systems with applications, 36(2),* 2009, 2592-2602.

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

[3]. Kim, S. Y., Jung, T. S., Suh, E. H., & Hwang, H. S, Customer segmentation and strategy development based on customer lifetime value: A case study, *Expert systems with applications, 31(1),* 2006, 101-107.

[4]. Hsieh, N. C, An integrated data mining and behavioral scoring model for analyzing bank customers, *Expert systems with applications, 27(4),* 2004, 623-633.

[5]. McCarty, J. A., & Hastak, M, Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression, *Journal of business research, 60(6),* 2007, 656-662.

[6]. Bloom, J. Z, Market segmentation: A neural network application, *Annals of Tourism Research, 32(1),* 2005, *93-111*

[7]. Huang, J. J., Tzeng, G. H., & Ong, C. S, Marketing segmentation using support vector clustering, *Expert systems with applications, 32(2),* 2007, 313-317.

[8]. Kuo, R. J., Ho, L. M., & Hu, C. M. Integration of self-organizing feature map and K-means algorithm for market segmentation, *Computers & Operations Research, 29(11),* 2002, 1475-1493.

[9]. Lee, J. H., & Park, S. C, Intelligent profitable customers segmentation system based on business intelligence tools, *Expert systems with applications, 29(1),* 2005, 145-152

[10]. Wu, R. S., & Chou, P. H, Customer segmentation of multiple category data in e-commerce using a soft-clustering approach, *Electronic Commerce Research and Applications, 10(3),* 2011, 331-341.

[11]. Tsai, C. Y., & Chiu, C. C, A purchase-based market segmentation methodology, *Expert Systems with Applications, 27(2),* 2004, 265-276.

[12]. Lu, T. C., & Wu, K. Y, A transaction pattern analysis system based on neural network, *Expert Systems with Applications, 36(3),* 2009, 6091-6099.

[13]. Hsu, F. M., Lu, L. P., & Lin, C. M, Segmenting customers by transaction data with concept hierarchy, *Expert Systems with Applications, 39(6),* 2012, 6221-6228.

[14]. Lu, K., & Furukawa, T, A Framework for Segmenting Customers Based on Probability Density of Transaction Data, *In Advanced Applied Informatics (IIAIAAI), 2012 IIAI International Conference on IEEE,* 2012, September, 273-278.

[15]. Oh, S. J., & Kim, J. Y, A hierarchical clustering algorithm for categorical sequence data, *Information Processing Letters, 91(3),* 2004, 135-140.

[16]. De Angelis, L., & Dias, J. G, Mining categorical sequences from data using a hybrid clustering method, *European Journal of Operational Research, 234(3),* 2014, 720-730.

[17]. Xu, W., Liu, X., & Gong, Y,. Document clustering based on non-negative matrix factorization, *In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval,* 2003, 267-273

[18]. Lee, J. H., Park, S., Ahn, C. M., & Kim, D, Automatic generic document summarization based on non-negative matrix factorization, *Information Processing & Management, 45(1),* 2009, 20-34.

[19]. Chen, Y. L., Kuo, M. H., Wu, S. Y., & Tang, K, Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data, *Electronic Commerce Research and Applications, 8(5),* 2009, 241-251.

[20]. Li, Y., & Ngom, A, The non-negative matrix factorization toolbox for biological data mining, *Source code for biology and medicine, 8(1),* 2013, 1