

# Mining Cluster-based Patterns for Elder Self-care Behavior

Yu-Shiang Hung<sup>1</sup> Kuei-Ling B. Chen<sup>2</sup> Chi-Ta Yang<sup>3</sup> Guang-Feng Deng<sup>4</sup>

Innovative DigiTech-Enabled Applications & Services Institute

Institute for Information Industry

8F., No.133, Minsheng E. Rd., Taipei City 105, Taiwan

garhung@iii.org.tw<sup>1</sup> klchen@iii.org.tw<sup>2</sup> gary1122@iii.org.tw<sup>3</sup> raymaldeng@iii.org.tw<sup>4</sup>

## Abstract

The rapid growth of the elderly population has increased the need to support elders in maintaining independent and healthy lifestyles in their homes rather than through more expensive and isolated care facilities. Self-care can improve the competence of elderly participants in managing their own health conditions without leaving home. This main purpose of this study is to understand the self-care behavior of elderly participants in a developed self-care service system that provides self-care service and to analyze the daily self-care activities and health status of elders who live at home alone.

To understand elder self-care patterns, log data from actual cases of elder self-care service were collected and analysed by Web usage mining. This study analysed 3391 sessions of 157 elders for the month of March, 2012. First, self-care use cycle, time, function numbers, and the depth and extent (range) of services were statistically analysed. Association rules were then used for data mining to find relationship between these functions of self-care behavior. Second, data from interest-based representation schemes were used to construct elder sessions. The *ART2*-enhance *K*-mean algorithm was then used to mine cluster patterns. The analysis results can be used for research in medicine, public health, nursing and psychology and for policy-making in the health care domain.

**Keywords:** Elder self-care behavior pattern; web usage mining; cluster analysis; sequential profiles; Markov model; association analysis.

## 1 Introduction

The global population is aging rapidly and is expected to require expanded health care services and facilities. However, since many elders prefer to stay in their private residences for as long as possible, methods are needed to enable them to do so safely and at reasonable costs. Several studies have investigated how information technologies can assist elders in independent living and in daily activity (Barger, Brown et al. 2005; Chung-Chih, Ming-Jang et al. 2006; Honda, Fukui et al. 2007; Leijdekkers, Gay et al. 2007; Seon-Woo, Yong-Joong et al. 2007; Wang, Wang et al. 2007; Dobrescu, Dobrescu et al. 2009)

.A major goal of elderly care is facilitating the ability of elders to maintain and promote their own health. Although they may suffer from chronic disease, cognitive impairment and functional limitation, mobilization of their self-care resources can minimize their health problems and enhance their health and well-being (Hoy, Wagner et al. 2007). Even when they have chronic disease and disability, elders often consider themselves active and in good health, and they are usually highly motivated to learn about ageing and health conditions.

Self-care can improve the competence of elders in managing their own health conditions without institutional care (Hoy, Wagner et al. 2007). However, a clear understanding of self-care activities in this group is essential (Arcury, Grzywacz et al. 2012). In 2011, the ComCare elder self-care project in Taiwan established an integrated IT platform that enables elders living at home to use tablet computers for major physical and mental health self-care functions. The ComCare server system has a database server and a web server for collecting useful information about health status and daily activities. The system enables elders to enter the information by using tablet PCs. Web Usage Mining is an area of Web Mining that deals with extracting interesting and useful knowledge from logging information produced by Web servers (Facca and Lanzì 2005; Sajid, Zafar et al. 2010; Wang and Lee 2011). Many researchers have applied Web usage mining for characterizing usage based on navigation patterns (Chen, Bhowmick et al. 2009; Bayir, Toroslu et al. 2012), for behavior prediction (Dimopoulos, Makris et al. 2010), for personalized recommendation (Mobasher, Cooley et al. 2000; Pierrakos, Paliouras et al. 2003; Park, Kim et al. 2012) and for web service improvement (Carmona, Ramírez-Gallego et al. 2012).

The main purpose of this study is to apply data mining techniques, including statistical analysis, clustering, association rules and sequential pattern discovery, for mining Web usage information from ComCare server logs to understand elder self-care behavior patterns. First, a statistical analysis is performed to analyze self-care use cycle, time, function numbers, and association rules to determine the depth and extent (range) of ComCare use. Second, interest-based representation schemes are used to construct elder sessions and then combined with an *ART2*-enhance *K*-mean algorithm to mine cluster patterns. To capture sequence information for self-care behavior patterns for elders using ComCare, sequence-based representation schemes in association with Markov models are combined with an *ART2*-enhance *K*-mean algorithm for mining cluster patterns in elder behavior. The analysis results can be used for research and policy-making by health care experts in medicine, public health, nursing and psychology. In practice, the improved

---

Copyright (c)2012, Australian Computer Society, Inc. This paper appeared at the 10th Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, December 2012. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 134, Yanchang Zhao, Jiuyong Li, Paul Kennedy, and Peter Christen, Ed. Reproduction for academic, not-for-profit purposes permitted provided this text is included.

characterization of elder self-care based on the analytical results in this study can improve personalized service and the design of elder self-care services.

This study is organized as follows Section 2 presents the functions of ComCare. Section 3 describes the Web usage mining technique used to analyse elder self-care in this study, including interest-based representation schemes, and ART2-enhance  $K$ -mean algorithm. Section 4 presents the experimental results. Section 5 presents conclusions and proposes future works.

## 2 ComCare Functions Description

The global population is aging rapidly. Taiwan was classified as an aging society in 1993, and will be classified as an aged society by 2018. The major concern is the speed of aging. According to Ministry of Interior data, 10.63% of the total Taiwan population were older than sixty-five years in 2010, and the aging index was 65.05%. The population is aging faster than any country in the world due to the low birth rate, which was only 0.9 in 2010.

ComCare elder self-care service was provided by *Institute for Information Industry Innovative DigiTech-Enabled Applications & Services Institute (IDEAS)* in Taiwan. The service requires an in-house tablet PC device and a server system. The server system includes a database server and a web server, which collect useful behavior data log from the tablet PC device. ComCare currently offers 14 self-care main-functions to address the everyday needs of seniors (Fig. 1); Of these, five are health management-related functions: ‘Diet Management’, ‘Exercise Management’, ‘Blood pressure Management’, ‘BMI Management’, and ‘Statistical data Management’. The three social community behavior-related main-functions are ‘Friend Management’, ‘Video Interaction’ and ‘Photo sharing’ in interactive care topic. The five life-information-related main-functions are ‘On sale’, ‘Calendar management’, ‘Weather forecast’, ‘Resident information’ and ‘Community management’ in life-information topic. ‘Entertainment management’ is the only entertainment-related main-function in the entertainment category. Additionally, each main function has various sub-functions; ComCare provides 71 such sub-functions for elder self-care service.



Figure 1: Main functions in elder care services

## 3. Methodology

This section first describes the data log preprocessing procedure and then presents two representation schemes suitable for capturing elder self-care behavior, including interest-based and sequence-based representation in Web

usage mining. Finally, the ART2 neural network and  $K$ -mean clustering algorithm used in this study are introduced.

### 3.1 Web-log Preprocessing

Data log preprocessing transforms the original logs so that all web access sessions can be identified. The Web server usually registers the access activities of website users in Web server logs. Different server parameters settings result in many different web log types, but log files typically share the same basic information, including client IP address, request time, requested *URL*, *HTTP* status code, referrer, *etc.* Generally, several preprocessing tasks are required before performing web usage mining algorithms on the Web server logs. The tasks in this work include data cleaning, user differentiation and session identification.

These preprocessing tasks resemble those in any other web usage mining problem and are discussed in detail in Hussain *et al.* (2010). (Hussain, Asghar *et al.* 2010) The original server logs are cleansed, formatted, and then grouped into meaningful sessions before use in web usage mining. A session can be described as the self-care activities performed by an elder between the start and the end of the ComCare session. Therefore, session identification is the process of segmenting the access log of each elder into individual access sessions. The activity stay time-based method developed by Liu and Kešelj (2007) was applied for session identification in this study (Liu and Kešelj 2007). This method limits the time spent on a function of ComCare to a specified threshold. If the time between the request most recently assigned to a session and the next request from the elder exceeds the threshold, a new access session is assumed. A 10 min activity-stay time is considered a conservative threshold for capturing the time for loading and studying page content (Liu and Kešelj 2007). However, this study uses the average use time of the function of all sessions as a threshold for function use time.

### 3.2 Interest-based Representation

‘Frequency’ and ‘Duration’ information for elder self-care behavior is captured by using an interest-based representation similar to that described in Lin, Liu and Kešelj (2007) (Liu and Kešelj 2007). Two indicators, ‘Frequency’ and ‘Duration’, are used to represent the interest sessions of elders. Let  $F$  be a set of sub-functions used by elders in ComCare server logs,  $F = \{f_1, f_2, \dots, f_m\}$ , each of which is uniquely represented by its associated sub-function ID. Let  $S$  be a set of user access sessions. Hence,  $S = \{s_1, s_2, \dots, s_n\}$ , where each  $s_i \in S$  is a subset of  $F$ . To facilitate the clustering operation, each session  $s$  is represented as an  $m$ -dimensional vector over the space of sub-functions,  $s = \{(int_1, s), (int_2, s), \dots, (int_m, s)\}$ , where  $(int_i, s)$  is an interest assigned to the  $i$ th sub-function ( $1 \leq i \leq m$ ) used in a session  $s$ . All sub-functions are assumed to be equally important to elder self-care pattern profiles. Therefore, regardless of the self-care sequence, the focus is the specific sub-functions used in a session.

The interest  $(int_i, s)$  must be determined appropriately to capture the interest of an elder in a sub-function. Two underlying concepts of this measure are ‘Frequency’ and

‘Duration’. ‘Frequency’ is the number of uses of a sub-function. A high frequency for a sub-function is assumed to indicate a strong need or interest of the elders. Equation (1) is the formula for ‘Frequency’, which is normalized by the total number of uses of sub-functions in the session:

$$Frequency(f_i) = \frac{NumberOfUses(f_i)}{\sum_{f_j \in UsedFunctions} NumberOfUses(f_j)} \quad (1)$$

‘Duration’ is defined as the time spent on a sub-function, *i.e.*, the difference between the requested time of two adjacent entries in a session. We conjecture that, as the time spent on a sub-function increases, the likelihood of the elder becoming interested in the sub-function increases. Typically, an elder quickly jumps to another sub-function if a sub-function is not useful. However, a quick jump might also result from the short operating length of a sub-function. Hence, normalizing ‘Duration’ by the operating length of the sub-function, that is, by the basic operating time of the sub-function, is more appropriate. Equation (2) is used to measure the ‘Duration’ of a sub-function,

$$Duration(f_i) = \frac{TotalDuration(f_i) / Length(f_i)}{\max_{f_j \in UsedFunctions} TotalDuration(f_j) / Length(f_j)} \quad (2)$$

where ‘Duration’ of a sub-function is further normalized by the max ‘Duration’ of sub-functions in the session. For the last access sub-function in each user access session, the duration cannot be estimated by calculating the difference in requested time. The average duration of the relevant session is used as the estimated duration for the last access event.

In this study, ‘Frequency’ and ‘Duration’ are considered two strong indicators of interest of elders. Therefore, ‘Frequency’ and ‘Duration’ are valued equally in the proposed interest measure. The following equation shows how the harmonic means for ‘Frequency’ and ‘Duration’ are used to represent the interest of an elder in a sub-function during a session:

$$Interest(f_i) = \frac{2 \times Frequency(f_i) \times Duration(f_i)}{Frequency(f_i) + Duration(f_i)} \quad (3)$$

Equation (3) ensures that ‘Interest’ for a sub-function is high only when both ‘Frequency’ and ‘Duration’ are high. Meanwhile, ‘Interest’ is normalized to a value between 0 and 1 which is convenient not only for understanding, but also for session clustering.

Each elder access session is eventually transformed into an  $m$ -dimensional vector of interests of sub-functions, *i.e.*,  $s = \{int_1, int_2, \dots, int_m\}$ , where  $m$  is the number of sub-functions used in all user access sessions. However, if the number of dimensions  $m$  exceeds a reasonable size, it not only consumes substantial processing time during clustering sessions, but also limits the real-world applicability of the system. Dimensions are reduced by using a frequency threshold  $f_{min}$  as a constraint to filter out sub-functions that are accessed less than  $f_{min}$  times in all access sessions. Our research showed that 60% of sub-functions appearing in the access sessions were visited at least 50 times. These sub-functions were considered representative functions that drew the attention of elders. Therefore,  $f_{min}$  was set to 50.

### 3.3 Clustering Analysis

In Web usage mining, clustering finds groups that share common properties and behavior by analyzing the data collected in web servers. Given the transformation of elder self-care access sessions into a multi-dimensional space as interest-based representation vectors or sequence-based representation matrices of functions, a clustering algorithm was applied to the derived elder self-care access sessions. Since access sessions are the images of activities by elders, representative elder self-care patterns can be obtained by clustering. These patterns also facilitate profiling of elder users of the ComCare service. This section describes how session clustering is performed and how cluster number is determined.

#### 3.3.1 Optimizing the Number of Clusters

Since the used clustering algorithm is a supervised clustering method, an *ART2* neural network (Kuo and Lin 2010) is needed to determine the number of clusters. The *ART2* neural network architecture is designed for processing both analog and binary input patterns. An *ART2* neural network consists of  $F_1$  and  $F_2$  layers. The  $F_1$  layer has seven nodes ( $W, X, U, V, P, Q$ ). The input signal is processed by the  $F_1$  layer and then passed from the bottom to the top value ( $b_{ij}$ ). The result of the bottom-to-top value is an input signal of the  $F_2$  layer. The nodes of the  $F_2$  layer compete with each other to produce a winning unit, which returns the signal to the  $F_1$  layer. The match value is then calculated with the top-to-bottom value ( $t_{ji}$ ) in the  $F_1$  layer and compared with the vigilance value. If the match value exceeds the vigilance value, then the weights of  $b_{ij}$  and  $t_{ji}$  are updated. Otherwise, the reset signal is sent to the  $F_2$  layer, and the winning unit is inhibited. After inhibition, the other winning unit is found in the  $F_2$  layer. If all  $F_2$  layer nodes are inhibited, the  $F_2$  layer produces a new node and generates the initial weights corresponding to the new node.

#### 3.3.2 Session clustering

After the *ART2* neural network determines the number of clusters, standard clustering algorithms can partition this space into groups of sessions that are close to each other based on a distance measure. The well-known *K*-means algorithm is used as the base method for clustering interest-based representation sessions and sequence-based representation sessions. The *K*-means clustering algorithm groups sessions by attributes/features into a  $k$  (positive integer) number of groups by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Additionally, the most popular Euclidean distance is used as the distance measure. The *K*-means clustering algorithm is performed in the following steps: Step 1: Generate initial random cluster centroids for  $k$  clusters and  $k$  obtained by *ART2* neural network. Step 2: Assign each session to its closest cluster centroid in terms of Euclidean distance. Step 3: Compute new cluster centroids. Step 4: If cluster memberships differ from the last iteration, repeat steps 2–3. Step 5: Stop and store clustering result.

Session clustering obtains a set of clusters,  $C = \{c_1, c_2, \dots, c_k\}$  in which each  $c_i$  ( $1 \leq i \leq k$ ) is a subset of the set of elder access sessions  $S$  where  $k$  is the number of clusters. A mean vector  $m_c$  of a sequence-based

representation (a mean matrix  $m_c$  of an interest-based representation) is computed as a representation for each session cluster  $c \in C$ . Each mean vector represents the representative user self-care pattern for a cluster in which a particular set of functions are accessed. The mean value for each function in the mean vector is computed as the average weight of the functions across total access sessions in the cluster. Therefore, the mean value is also between 0 and 1. Meanwhile, a weight threshold for the mean vector of each session cluster,  $w_{min}$ , is set as a constraint to filter out functions with mean values below the threshold for the cluster. The remaining self-care functions in each cluster are considered of greatest interest to elders and are used as representative self-care patterns for the cluster. Since the least mean value is always far smaller than the second least and the third least mean values, the second least mean value of each mean vector is used as the  $w_{min}$  for each session cluster.

In our research elder self-care patterns are described in terms of the common usage characteristics for a group of elders. Since many elders may have common interests up to a point during their self-care navigation, navigation patterns should capture the overlapping interests or the information needs of these users. In addition, self-care patterns should also be capable to distinguish among functions based on their different significance to each pattern. This work defines an elder self-care pattern  $np$  as a pattern that captures an aggregate view of the behavior of a group of elders based on their common interests or sequence information. After session clustering,  $NP = \{np_1, np_2, \dots, np_k\}$  represents the set of elder self-care patterns, in which each  $np_i$  is a subset of  $F$ , the set of functions.

#### 4. Elder Self-care Behavioral Analysis

##### 4.1 Dataset and Environment

The ComCare data server was used to obtain a record of each function used by an elder from the start of service until the present. For all of March, 2012, 3391 sessions of 157 elders were identified for analysis. Matlab Language is used to perform the web usage mining algorithm. In addition to collecting the record data, the researchers invited ComCare users to complete a questionnaire survey, which recovered 157 valid questionnaires. The content of the structured questionnaire, which was compiled by a team of experts, includes user impressions of ComCare service (including information quality, service quality, ComCare self-efficacy, and perceived risk. To prevent users from giving noncommittal responses and to ensure easily measurable user responses, a six-item Likert scale with six choices for each dimension was used; the higher the score for each item, the greater the satisfaction of the respondent with that service. Analysis focused on the record data and questionnaire data for the 157 test elders who completed the questionnaire.

The questionnaire results indicated that the age range of the test users was 58 to 86 years, and the average age was 68 years (standard deviation, 8.22). Figure 2 presents basic information concerning ComCare users. Of the 157 test users, 56% were women and 44% were men. Education levels were relatively high; most had at least a university education. At least 75% of the respondents regularly used

the Internet, which was higher than the finding of another survey of internet use among the older generation, which showed that 56.3% of senior respondents had internet experience. Approximately 68% of users were retired or not currently employed, and close to 40% of users had a history of chronic disease such as hypertension or diabetes.

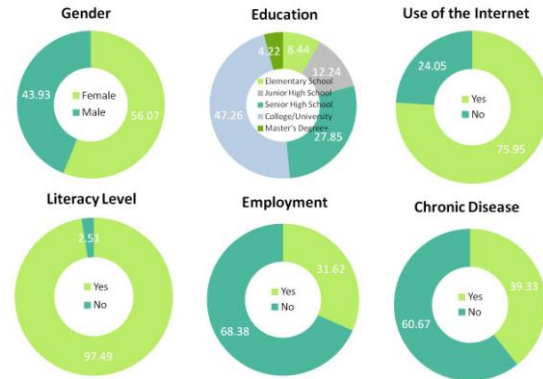


Figure 2: Basic background information of self-care elders

##### 4.2 Elder self-care Profiling Results

Two factors were considered when profiling the self-care behavior of elders: use function length in terms of the number of sub-functions used by elders in ComCare and use duration in terms of the number of hours spent by elders in ComCare. Figure 3 shows the distribution of use function length for elder self-care. The mean number of elder requests per session is almost 4.5, which is lower than the expected number of the ComCare provider.

Comparison with other statistics indicated that the distribution of use functions is strongly right-skewed. For example, the mean (4.5) is higher than the median (3), the mode (1) is the same as the minimum, and the maximum (24) is very large. Figure 3 confirms our deduction that the distribution is right-skewed (to increase granularity, use functions above 20 are omitted from the graph. Inclusion of these records would have made the graph even more right-skewed). The data shows that the great majority of sessions included fewer than five function requests, half included three or fewer, and a disappointingly large number of sessions included only a single function request. This finding should be noted by service developers because it raises the question of why elders are leaving so soon. Notably, however, the third and fourth most common actions included five and eight actions, respectively, which is higher than the expected number.

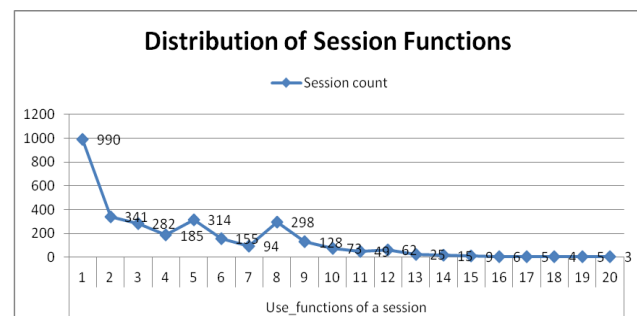


Figure 3: Distribution of session functions



Apart from the number of function requests per session, another important variable is the duration of time per session that the elders spent on the ComCare self-care service. The mean session duration is 554.8 seconds (9.23 minutes). Again, the outcome is lower than the expected number of the ComCare provider. However, the median for right-skewed data (Fig. 4) is a better summary statistic than the mean. The median session duration is 318.5 seconds (about 5.28 minutes), which is a more realistic estimate of the typical duration of a session for those who requested more than one function. Figure 4 shows the distribution of session duration (also known as self-care use time) for multi-function sessions. Again, the upper tail is clipped at 2400 seconds to increase the granularity of the graph. The figure shows that most sessions lasted less than 10 minutes, half lasted 6 minutes or fewer, and a disappointingly large number of sessions lasted only used 3 minutes. Again, service developers should be concerned.

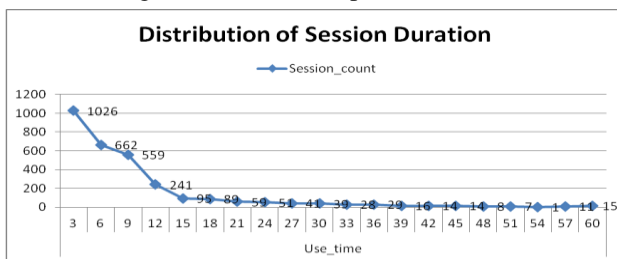


Figure 4: Distribution of session duration

Hierarchical analysis was also performed to understand the depth and extent of use of ComCare. Figure 5 shows the hierarchical architecture of ComCare, which shows that the health management-related sub-functions were the most frequently used activities and that each sub-function of these health topics had a strong association.

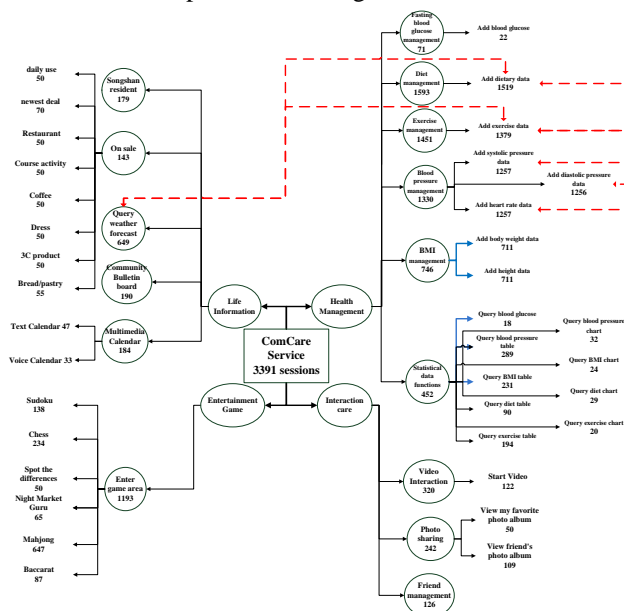


Figure 5: Depth and extent of self-care service use

Among the four topics, the use depth of interaction care and the life information topic revealed low use level. Only about 5% of the total usage session was spent on the sub-function of the Life information topic, excluding Query weather forecast, and only about 30% of usage session was spent on the main-function of the Life information topic keep going to request the

main-function’s sub-functions of Life information. The interaction care topic reveals the same condition. The main-function of the interaction care topic was visited in only about 8% of the total usage sessions, and, of these, the sub-function was again used in only about 30%. The results indicate that self-care service providers need further information to improve the low rate of self-care functions.

### 4.3 Association Analysis of ComCare Service

This paper applied an *a priori* algorithm to the self-care service log data to reveal actionable association rules. The condition support > 5% and confidence > 90% revealed 63 association rules. For brevity, Table 1 lists only three interesting association rules revealed by the *a priori* algorithm. The rules are sorted by “rule support”. Consider the first rule in the list. The antecedent is Exercise management function, and the consequent is Diet management function. The form of the rule is therefore Exercise management ( 1376 ) ⇒ Diet management( 1265 ) conf:(0.92). The support of the rule is 37.46%, meaning that the rule applies to almost 37% of the 3391 total sessions in the data, which is a very high support level. A 92% confidence indicated that, of these 3391sessions, 1376 met the antecedent condition; that is, 1376 requested Exercise management at some point. Of these 1376 sessions, the Diet management function was also requested in 92% (1265 sessions). According to rule 1, elders who pay attention to the Exercise management also emphasize Diet management in self-care behavior. According to rule 2, elders who pay attention to the BMI management also emphasize Blood management in self-care behavior. Rule 3 is that elders who pay attention to the BMI management & Exercise management also emphasize Diet management & Blood management. Surprisingly, in the support > 5% and confidence>90% condition, the analysis showed that, of the functions in the four different topic, only the Query weather forecast function in the life-information-related topic was associated with Diet management and Exercise management in the health management-related topic.

Table 1: Association analysis of self-care behavior

Rule ID	Antecedent(account) ⇒ Consequent(account)	(Sup,Conf)
1	Exercise management ( 1376 ) ⇒ Diet management ( 1265 )	(0.37,0.92)
2	BMI management ( 708 ) ⇒ Blood management ( 593 )	(0.18,0.84)
3	BMI management & Exercise management (659) ⇒ Diet management & Blood management ( 556 )	(0.16,0.84)

### 4.4 Interest-based Cluster Patterns

This section describes the elder self-care pattern obtained in three clusters by applying interest-based representation schemes and ART2-enhance K-mean algorithm. Table 2 characterizes each cluster pattern. Based on the observed means and proportions in Table 2, cluster C1 can be labelled as “Health dominate elders”. The BMI, Diet, Exercise, and Blood pressure management functions were

requested at a much higher rate and for a longer duration by Health elders compared to elders in other clusters. Cluster C2 can be labelled as “Entertainment dominate elders”. The Mahjong game functions are requested by these elders at a rate thousands of times higher compared to other typical elders. The same was true for the chess game. Cluster C3 can be labelled “General elders”. The Diet and Exercise management functions were requested at a much higher rate and for a longer duration compared to the other functions. Cluster C3, which is the largest cluster, contains 59 elders; cluster C1 contains 52 elders; cluster C2 contains 46 elders. These data do not enable provisional identification of the “typical” cluster because the three clusters comprise 30%~40% of the sessions.

Health-dominant elders use more self-care functions (8.46) compared to Entertainment dominate elders (2.74) and General dominate elders (4.1) although the session duration of Health dominate elders is the shortest among the three cluster types since the average time per function is a fraction of that of the Entertainment dominate elders. The use cycle of General dominate elders is longest among the three cluster types. The Health dominate elders, however, very rarely access game functions. For all cluster types, Table 2 also shows no interest in functions of Friend management, Photo sharing, Multimedia calendar, Video Interaction and Enter community bulletin board functions. This raises the questions of whether these self-care functions are sufficiently user friendly and how the functions can be changed to induce elders to linger and use more functions. Thus, the interest-based representation scheme reveals interesting differences between the three elder self-care patterns. Perhaps the elder self-care service masters could apply this knowledge to differentiate service for each of the three elder types.

**Table 2: Elder self-care characterization of interest-based cluster**

Function	C1 Health-dom inate elders	C2 Entertainme nt-dominate elders	C3 General elders
People Account	52	46	59
percentage	33.12%	29.30%	37.58%
Cycle ( day )	1.511	1.444	1.845
session duration	382.636	657.045	409.553
Function number	8.467	2.742	4.07
Friend management	0.055	0.048	0.022
BMI management	0.851*	0.026	0.06
Diet management	0.865*	0.164	0.878*
Exercise management	0.878*	0.122	0.758*
Photo sharing	0.034	0.04	0.037
Calendar Management	0.04	0.049	0.051
Video Interaction	0.115	0.103	0.101
Blood management	0.785*	0.291	0.447

Weather forecast	0.319	0.262	0.079
Enter community bulletin board	0.079	0.062	0.037
Game - Chess	0.014	0.111*	0.011
Game - Mahjong	0.074	0.354*	0.106

\*: High-interest function

## 5 Conclusion and Future Work

To improve understanding of the self-care behavior of elders living alone, this study analysed real-world data for elder self-care service by applying Web usage mining methodology, including association analysis, and interest- and sequence-based representation schemes in association with Markov models combined with ART2-enhance K-mean algorithm.

The analysis results show that elder self-care behavior can be classified by an interest-based representation scheme into three distinct cluster types: health-dominate type, entertainment-dominate type and general-dominate type. Each type displays different self-care use cycles, times, function numbers and needs. However, six distinct cluster types can be identified by sequence-based clustering from different use sequence. Each type provides detailed information about self-care use cycle, time, function numbers and characterizations. This research shows that the use of sequence-based clustering in web usage mining effectively finds meaning groups that share common interests and behaviors and effectively extracts knowledge needed to understand the motivation for using elder self-care. The analysis results can be used by experts in medicine, public health, nursing and psychology to further research and to assist in policy-making in the health care domain. Future research will apply the sequence-based clustering results for the ComCare project to improving personalized elder self-care services.

## 6 Acknowledgments

This study is conducted under the “Smart Living Technology and Service program” of the Institute for Information Industry which is financially supported by the Ministry of Economy Affairs of the Republic of China.

## 7 References

- Arcury, T. A., J. G. Grzywacz, et al. (2012): Older adults' self-management of daily symptoms: Complementary therapies, self-care, and medical care. *Journal of Aging and Health* **24**(4): 569-597.
- Barger, T. S., D. E. Brown, et al. (2005): Health-status monitoring through analysis of behavioral patterns. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* **35**(1): 22-27.
- Bayir, M. A., I. H. Toroslu, et al. (2012): Discovering better navigation sequences for the session construction problem. *Data & Knowledge Engineering* **73**(0): 58-72.
- Carmona, C. J., S. Ramírez-Gallego, et al. (2012): Web usage mining to improve the design of an e-commerce website: OrOliveSur.com. *Expert Systems with Applications* **39**(12): 11243-11249.

- Chen, L., S. S. Bhowmick, et al. (2009): COWES: Web user clustering based on evolutionary web sessions. *Data & Knowledge Engineering* **68**(10): 867-885.
- Chung-Chih, L., C. Ming-Jang, et al. (2006): Wireless health care service system for elderly with dementia. *Information Technology in Biomedicine, IEEE Transactions on* **10**(4): 696-704.
- Dimopoulos, C., C. Makris, et al. (2010): A web page usage prediction scheme using sequence indexing and clustering techniques. *Data & Knowledge Engineering* **69**(4): 371-382.
- Dobrescu, R., M. Dobrescu, et al. (2009): Embedded wireless homecare monitoring system. *eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED '09. International Conference on.*
- Facca, F. M. and P. L. Lanzi (2005): Mining interesting knowledge from weblogs: a survey. *Data & Knowledge Engineering* **53**(3): 225-241.
- Honda, S., K. I. Fukui, et al. (2007): Extracting human behaviors with infrared sensor network. *Networked Sensing Systems, 2007. INSS '07. Fourth International Conference on.*
- Hoy, B., L. Wagner, et al. (2007): Self-care as a health resource of elders: an integrative review of the concept. *Scandinavian Journal of Caring Sciences* **21**(4): 456-466.
- Hussain, T., S. Asghar, et al. (2010): Web usage mining: A survey on preprocessing of web log file. *Information and Emerging Technologies (ICIET), 2010 International Conference on.*
- Kuo, R. J. and L. M. Lin (2010): Application of a hybrid of genetic algorithm and particle swarm optimization algorithm for order clustering. *Decision Support Systems* **49**(4): 451-462.
- Leijdekkers, P., V. Gay, et al. (2007): Smart homecare system for health tele-monitoring. *Digital Society, 2007. ICDS '07. First International Conference on the.*
- Liu, H. and V. Kešelj (2007): Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data & Knowledge Engineering* **61**(2): 304-330.
- Mobasher, B., R. Cooley, et al. (2000): Automatic personalization based on Web usage mining. *Communications of the ACM* **43**(8): 142-151.
- Park, D. H., H. K. Kim, et al. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications* **39**(11): 10059-10072.
- Park, S., N. C. Suresh, et al. (2008): Sequence-based clustering for Web usage mining: A new experimental framework and ANN-enhanced K-means algorithm. *Data & Knowledge Engineering* **65**(3): 512-543.
- Pierrakos, D., G. Paliouras, et al. (2003): Web Usage Mining as a Tool for Personalization: A Survey. *User Modeling and User - Adapted Interaction* **13**(4): 311-372.
- Sajid, N. A., S. Zafar, et al. (2010): Sequential pattern finding: A survey. *Information and Emerging Technologies (ICIET), 2010 International Conference on.*
- Seon-Woo, L., K. Yong-Joong, et al. (2007): A remote behavioral monitoring system for elders living alone. *Control, Automation and Systems, 2007. ICCAS '07. International Conference on.*
- Wang, C. C. C., W. Y. Wang, et al. (2007): *Development of a service-oriented healthcare architecture for institution-based care. TENCON 2007 - 2007 IEEE Region 10 Conference.*
- Wang, Y.-T. and A. J. T. Lee (2011): Mining Web navigation patterns with a path traversal graph. *Expert Systems with Applications* **38**(6): 7112-7122.

