H-027 A Health Life-style Checke List and Analysis

Yohtaro Miyanishi†, Goutam Chakraborty‡, Koh Mizuno*
Mitsuaki Yamamoto* Atsushi Togashi†, Shoich Noguchi††

†: {miyanisi,togashi}@myu.ac.jp, Department of Design Information, Miyagi University,
‡: goutam@soft.iwate-pu.ac.jp, Department of Software and Information Science, Iwate Prefecture University,
*: {mizuno-k@tfu-mail.tfu.ac.jp, mituaki@ecei.tohoku.ac.jp}, Kansei Fukushi Research Center, Tohoku Fukushi University,
††: noguchi@sfais.or.jp, Sendai Foundation of Applied Information Science.

I. Introduction

In advanced countries, where the society is aging and the birth rate is dropping, health-related issues are most important to ensure high quality of life for individual citizens and to manage national medical budget. Taking those issues into account, a questionnaire named ”Healthy Life-style Check List” has been developed. ”Healthy Life-style Check List” [2], [3], [4] includes 98 questions which are categorized into 9 sections consisting of eating, respiration, sleep, thermoregulation, exercise, mind, sex, environment, and physical discomfort. In each section, there are 6 to 17 questions asking the behavior and physical and mental function/status which are followed by a question asking satisfaction level of the respective section. Therefore, the responders are led to think about their lifestyle, health status, and satisfaction of each section. Besides, as the questions to ask the behavior are based on scientific evidence beneficial for improving health, responders can check and mend their unhealthy habits.

Since the ”Healthy Life-style Check List” includes a variety of knowledge beneficial for health, it can be useful for health education. From this point of view, ”Healthy Life-style Check List” was applied to Japanese college students. Among college students, lack of healthy habits is prevalent, but at the same time, they tend to recognize their health as good. In the collected data of more than 4,000 of Japanese college students, possible clusters in answering the ”Healthy Life-style Check List” were investigated by using multivariate analysis.

II. Healthy Life-Style Check List

”Healthy Life-Style Check List (Genki Tenken Hyou in Japanese)” [2], [3], [4] is a health check questionnaire system which has been developed in the Kansei Fukushi Research Center of Tohoku Fukushi University, Sendai, Japan. A number of researchers in various scientific fields have supported to develop ”Healthy Life-Style Check List”. The questions are classified into nine sections and one over-all satisfaction level, including physical factors, spiritual factors and environmental factors. Each section reflects an aspect or mode of a person’s lifestyle. Those sections are (1)eating, (2)respiration, (3)sleep, (4)thermoregulation, (5)exercise, (6)mind, (7)sex, (8)environment, (9)physical discomfort, and (10)over-all satisfaction. The total number of questions is 98.

A person answers those 98 questions rating from one to five according to his/her daily lifestyle. Each ratings means 5: completely agreed, 4:almost agreed, 3:somewhat agreed, 2:somewhat disagreed, and 1:completely disagreed with the description of each question. Those questions have been refined on scientific reasoning and basis. Answering those questions, he/she can check his/her current lifestyle and can improve it. The questions are listed in the table I.

TABLE I
Healthy Life-Style Check List

	(1) Eating
1)	Eat slowly
2)	Eat regularly three times a day
3)	Think about nutrition and safety
	...
17)	Regular urination
18)	Satisfaction about eating
	(2) Respiration (Breathing)
19)	Deep breathing consciously
20)	Do not smoke
	...
24)	No snoring
25)	Satisfaction about respiration
	(3) Sleep
26)	Wake up regularly
27)	Outdoor during daytime
	...
33)	No sleepiness during daytime
34)	Satisfaction about sleep
	(4) Thermoregulation
35)	Bath slowly in not too hot water
36)	Sunbathe during daytime
	...
41)	Bath comfortably
42)	Satisfaction about thermoregulation
	(5) Exercise
43)	Regular participation in some sports or dancing
44)	Bicycling
	...
50)	Sweating comfortably
51)	Satisfaction about exercise
	(6) Mind
52)	Future dream
53)	Life with objectives
	...
60)	Remembrance
61)	Satisfaction about mind
	(7) Sex
62)	Consideration to family and friends
63)	Love from family and friends
	...
70)	Talk about love in family
71)	Satisfaction about sex
	(8) Environment
72)	Cleanliness inside and outside of residence
73)	Classify garbage
	...
83)	Secure and peace
84)	Satisfaction about environment
	(9) Physical discomfort
85)	Have no headache
86)	Have no ringing in the ears
	...
96)	Appropriate engagement at back teeth
97)	Satisfaction about physical discomfort
	(10) Over-all satisfaction
98)	Over-all satisfaction level

III. Statistical Clustering of Healthy Life-Style Check List

The data collection was conducted after appropriate instructions of ”Healthy Life-Style Check List” and the purpose of planned statistical analysis. Then, the answers were gathered from 4036 college students who agreed to use their data for statistical analysis. Each answer consists of 98 items of value,

varying from 1 to 5. Though each student should answer all 98 questions, some students did not answer to every question.

We did different statistical analysis of the data to find the distribution pattern of life-style replies, itemwise, etc. We have gathered answers of above questionnaires from 4036 students. Each answer consists of 98 items of value, varying from 1 to 5. Though each student should answer all 98 questions, some students did not answer to every question.

The final objective of clustering analysis in this paper is to provide some useful advices to students to improve their lifestyle to become healthier. To realize this purpose, at first we should grasp the patterns of students' lifestyle i.e. classify the above answers into these patterns. And secondary extract characteristics of each pattern. And then we could provide useful advices. We analyzed the data as follows.

A. Preliminary Processing

1) Classify questions in ten sections to three type of questions and average value of answers: Clustering using whole 98 questions is not rather proper. We have experienced the fact. The reasons are:

At first, each section has different number of answers according to different number of questions. Next, some sections have similar questions in own section. And thirdly, some questions are not certain whether it gets better grading when answered logical AND type or logical OR type. These factors actually give unintended weights in clustering processing.

So we calculated the average values of three aspects of each section. The three aspects are behavior, status, and satisfaction.

We show the average processing in table II.

TABLE II
AVERAGE CALCULATION OF ANSWERS IN EACH SECTION

section	behavior	status	satisfaction
(1)	1)-11),13),14)	12), 15)-17)	18)
(2)	19)-21)	22)-24)	25)
(3)	26)-30)	31)-33)	34)
(4)	35)-39)	40)-41)	42)
(5)	43)-48)	49)-50)	51)
(6)	52)-56)	57)-60)	61)
(7)	62),70)	63)-69)	71)
(8)	72)-78)	79)-83)	84)
(9)	-	85)-96)	97)
(10)	-	-	98)

In the table, a)-b) means the average from a) to b), c),d) means the weighted average of c) and d).

After this processing, 98 answers (each value is from 1 to 5) are reduced into 27 answers (each value is also from 1 to 5)

2) Append ID to each student: In this analysis ID is indispensable. We appended ID each raw of original data table in automatic increment type.

3) Deletion of data including blank: As mentioned above, some students did not answer to some questions. Those data could not be clustered, so we deleted. Through this processing 4036 data reduces to 3596 data.

B. Clustering processing

The number of students (raw size) is 3596 and the number of answers (column size) is 27. This table was clustered and we got the dendrogram as result.

We use a free program [5] because we could not process a certain type of processing by SPSS. In this program, we choose parameters as follows.

1) Normalize data is (No normalization):

2) The distance of dendrogram is (Squared distance):

3) Clustering method is (Ward method): Ward method is known as a popular clustering method in multivariable statistical analysis.

C. Decision of the number of clusters

We need to select the number of clusters in advance. Too many clusters are difficult to handle to provide useful advices. Too few clusters do not make sense. We selected seven, viewing total tree form of resulting dendrogram.

SPSS gives us a dendrogram using graphical characters as oldfashioned semi-graphical displays. It does not use vector graphics. We analyze 3596 data, so it become very long diagram and we could not grasp whole form as it is.

In order to grasp the whole form, we must reduce longitudinal size of the result in dendrogram. It is essential to use vector graphics. The program1) provides vector graphics. We could reduce the size as shown in Fig. 1 . And then we selected the number of clusters as seven.

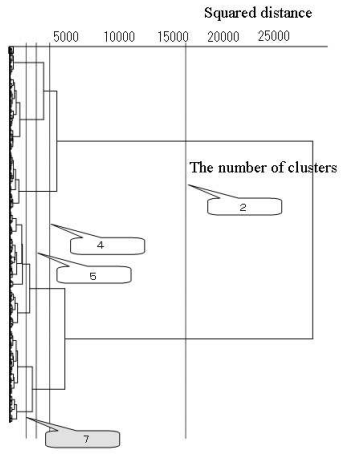

Fig. 1. Dendrogram

D. The average value of seven clusters

As mentioned above, we have classified 3596 data into seven clusters. We averaged three aspects (i.e. behavior, status, and satisfaction) of those students in same clusters. That is we calculated so called the center of gravity of each cluster. Those values are shown in Fig 2 represents the radar chart consisting of the averaged values of 8 sections ((1) eating to (8) environment) with the bar chart of section (9) physical discomfort and (10) over-all satisfaction. By using this form of representation, comprehensive characteristics of life-style, health status, and its satisfaction levels in each cluster can be observed. Cluster 1 and 2 show high values in every section, but the values are slightly less in cluster 2. Those students seemed to be good in their life-style and health status. In cluster 3, although the values are almost the same as those in cluster 2, behavioral and satisfaction levels in section (5) exercise are lower than 3 (2.59 and 2.33, respectively). As those students seemed to conduct less amount of daily exercise, detrimental consequences induced by inactivity had not appeared yet due to their physiological reservation of young age. Cluster 4 is characterized by the lowest values in every section in seven clusters, so that possible educational or behavioral interventions to improve their life-style are desired.

Cluster 5, 6 and 7 represent intermediate values with slightly different shape of radar chart among the clusters. The sections tend to show lower values are (1) eating (cluster 5 and 7), (3) sleep (cluster 5 and 7), (5) exercise (cluster 5 and 6), and (10) over-all satisfaction (cluster 5). Since eating, sleep and exercise are the essential factors for maintaining and improving health, possible interventions should be applied to the students of cluster 5 who showed lower values in all of three sections.

As this study was conducted based on the data collected from college students (mean age: 20.4 yrs), the difference in the values among seven clusters may be relatively small. In case of the data from middle to older aged population, individual difference in health status shall be larger and distinct feature of each cluster may appear. Further study is required on this point.

section	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cluster 1										
Behavior	3.75	3.94	4.32	4.03	3.91	4.43	3.76	4.25		
Status	4.32	4.46	4.12	4.67	4.35	4.56	4.50	4.50	4.64	
Satisfaction	4.61	4.80	4.53	4.75	4.39	4.70	4.42	4.68	4.71	4.54
Cluster 2										
Behavior	3.25	3.35	3.84	3.48	3.58	3.84	3.33	3.64		
Status	3.96	4.02	3.43	4.26	4.17	3.95	4.25	3.99	4.36	
Satisfaction	4.01	4.38	3.60	4.15	4.15	3.88	3.85	3.93	4.23	3.86
Cluster 3										
Behavior	3.26	3.33	3.94	3.47	2.59	3.86	3.37	3.61		
Status	3.93	4.35	3.77	4.48	3.29	4.10	4.29	4.15	4.42	
Satisfaction	3.87	4.58	3.95	4.35	2.33	4.04	3.88	4.06	4.23	3.64
Cluster 4										
Behavior	2.69	2.99	3.10	2.84	2.48	3.07	2.74	2.98		
Status	3.31	3.51	2.67	3.72	2.77	2.96	3.72	3.26	3.51	
Satisfaction	2.91	3.44	2.42	3.19	2.25	2.41	2.80	3.01	2.64	2.50
Cluster 5										
Behavior	2.53	2.62	3.22	2.80	2.26	3.14	2.88	3.02		
Status	3.44	4.12	2.98	4.07	2.95	3.29	3.99	3.61	4.03	
Satisfaction	2.46	4.14	2.63	3.52	1.95	2.85	3.37	3.45	3.40	2.71
Cluster 6										
Behavior	3.06	3.23	3.59	3.15	2.65	3.40	3.04	3.33		
Status	3.62	4.09	3.14	4.07	2.94	3.51	4.00	3.60	4.03	
Satisfaction	3.83	4.35	3.11	3.84	2.55	3.13	3.31	3.43	3.63	3.28
Cluster 7										
Behavior	2.83	3.03	3.36	3.23	3.48	3.55	3.22	3.32		
Status	3.59	3.62	2.92	4.15	4.03	3.43	4.09	3.62	3.71	
Satisfaction	3.02	3.72	2.74	3.73	3.92	3.16	3.49	3.44	3.20	3.00

Fig. 2. Average value of answers in each cluster and section

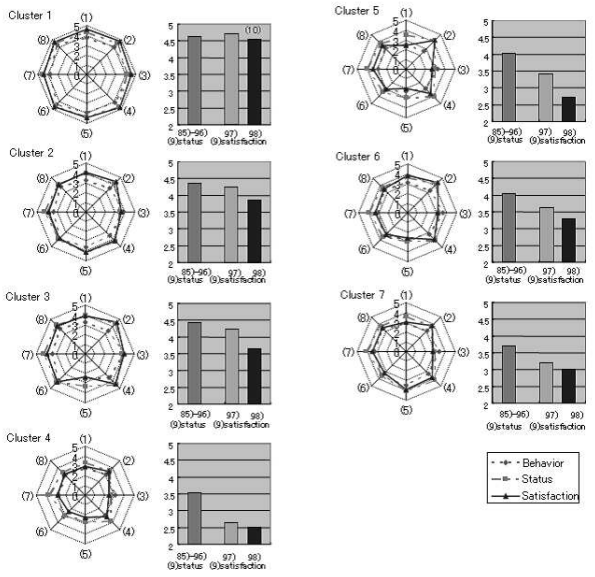

Fig. 3. Radar chart of seven clusters

IV. Clustering of the Life-style data Using Self Organizing Map

In the next step of data analysis, we used self organizing map (SOM) to visualize the data distribution and cluster formation if any. In this section, we will first give a brief introduction of self organizing map, and then the experments and results obtained using SOM.

A. Basic Operating Principle of SOM

An artificial neural network implementation of self organizing map and its training algorithm was first proposed by T. Kohonen [1]. It is an unsupervised learning algorithm by which a multi-dimensional data is projected onto a two dimensional plane for visualization of its distribution, including clusters if present in the original space.

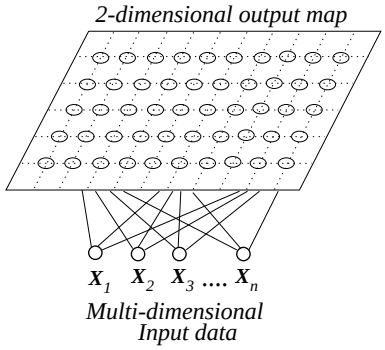

Fig. 4. Working principle of self organizing map

The multi-dimensional data, shown as $[X_1, X_2, \ldots X_n]$ in Fig. 4, are input to the network. For n-dimensional data, all the inputs are connected to all the outputs. Outputs are labeled in an ordered 2-dimensional matrix form, where the labels denote their position in the 2-dimensional output plane.

At the beginning, all connection weights between the input units and the output units are set to values $0.5 \pm \epsilon$, where ϵ is a small quantity. As evident from Fig. 4, all output units have n connection weights. Thus, every output unit represent a point in the original $n-$dimension data space. During training, while a data is input, the output unit whose position is nearest to the data, in the original $n-$dimension space, is the winner. That particular output unit's position is shifted a little towards the data, i.e., the connection weights to the winner output unit are modified so that it moves a little nearer to that input training data, in the original data space. Output units which are neighbors of the winner, as pre-assigned by the labels, also get trained but to a lesser extent. They also move towards the training sample like the winner, but by a still smaller value. Depending on the implementation, the neighbors could be the nearest 4 nodes, or may include the next ring of 8 surrounding nodes. The training intensity (the amount of movement of the node position due to training) decreases as we move further away from the winner node. The training continues for all the data for several times until convergence.

B. Results of SOM Experiments

The data set used is already explained in previous section. In this experiment we used only the final summary data of each item, namely food-habit, breathing, vision, feeling about temperature, exercise habits, mental status, nature and behavior, environment, and finally feeling of stability and security. In the preprocessing stage, we deleted incomplete data and finally had 3962 samples. We performed two sets of experiments, one with the first 1000 data, and the other

with the whole set of 3962 data. This is to check whether the behavior of the data is uniform over the whole data set or not. In both the cases SOM is run 2000 epochs, after which the training was stabilized. The results with 1000 data and 3962 data are shown in Fig. 5

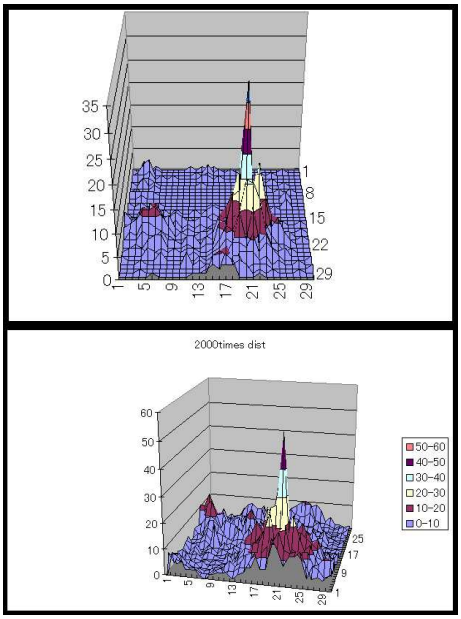

Fig. 5. Clustering of the first 1000 samples, and the whole set of 3962 samples by SOM

From the above figure it is observed that the formation of the cluster is similar for the first 1000 data as well as the whole set of data, ensuring that the behavior of the data is uniform over the whole set. Two prominent cluster centers, i.e., output nodes winning maximum number of data compared to its neighbors, were identified. Those output nodes are (5,20) and (20,2). The most prominent one is at (5,20). We collected data from those two cluster centers and the first few of them are shown in Table. III and Table. IV.

TABLE III
THE DETAIL OF A FEW DATA FROM THE LARGER CLUSTER

Data Sl.No.	food	breathe	vision	temp	exercise	mind	nature	environ	stability
28	5	5	5	5	5	5	5	5	5
75	5	5	5	5	5	4	5	5	5
129	5	5	5	5	5	5	5	5	5
149	5	5	5	5	5	5	4	5	5
221	5	5	4	5	5	5	4	5	5
281	5	5	5	5	5	5	5	5	5
284	5	5	4	5	4	5	5	5	5
309	5	5	4	5	5	5	5	5	5
344	5	5	5	5	5	5	5	5	5

TABLE IV
THE DETAIL OF A FEW DATA FROM THE SECOND LARGE CLUSTER

Data Sl.No.	food	breathe	vision	temp	exercise	mind	nature	environ	stability
1502	4	5	2	4	2	3	4	4	5
1608	4	5	3	4	2	3	3	5	5
1632	5	5	3	4	3	3	4	5	5
1774	4	5	3	4	2	3	4	4	5
1841	1	5	2	4	2	3	5	5	5
1944	5	4	3	5	2	4	3	4	5
2141	4	5	3	4	3	3	4	4	5
2148	4	5	3	5	3	3	3	4	4
2343	4	5	3	4	1	4	3	4	5

C. Analysis of the SOM result

Possible interpretation on the results of SOM analysis is stated in this subsection. The most prominent cluster seems to be identical with the cluster 1 found by cluster analysis who are satisfied with their own health status. Such tendency has been confirmed by previous studies conducted on young adults [6], [7] et al. As for the secondly prominent cluster showing lower satisfaction levels in sections of sleep, exercise and mind, students in this cluster might be overlapped with the cluster 3 to 7 found by cluster analysis. According to recent epidemiological studies on Japanese general population [8], [9], a significant relationship among those three items (sleep, exercise and mind) has been reported. As causal relationship among those items is unknown in the present study, further study to establish effective and efficient intervention procedure in relation to any or all of those items is desired.

V. DISCUSSION AND FUTURE WORK

In this experiment the number of items to be filled out is too many, making the process of data collection quite time consuming. One way of reducing the number of items is to find relationship among the items, and reduce the irrelevant ones. The best way to do that is to determine the eigen-values and principal component. Initially when principal component analysis is run over the whole data set, clear principal components were not clear. In the next step of experiment we are planning to collect data only from the cluster center and then use them for principal component analysis. We hope to find some correlation among items then.

One interesting extension of this work is to find a relation between life-style data and mental as well as physical health condition. If those relations or rules could be extracted, they can be used as directives.

At present, all the samples belong to same age group. Thus variations of data is small, and we have two prominant clusters only. In future, we will collect data from all members of the society to have more meaningful and versatile data set. The collection of such data itself has many difficult steps to overcome -like security, privacy issues. We are now working on implementing an internet environment to be able to collect such data.

ACKNOWLEDGMENT

”This work was carried out in the Kansei Fukushi Research Center of Tohoku Fukushi University, Sendai, Japan, supported by”Academic Frontiers” Project for Private Universities : matching fund subsidy from the Ministry of Education, Culture, Sports, Science and Technology (2004 ∼2008).” The work is also financially supported by the Strategic Information and Communications R and D Promotion Programme by the Ministry of Internal Affairs and Communication.

REFERENCES

[1] Teuvo Kohonen : *Self-Organization and Associative Memory*. Springer-Verlag.
[2] M. Yamamoto, K. Mizuno, et al., An Approach of the Prevention Welfare Health Center in Tohoku Fukushi University, – Toward the Establishment of Self Health Management with SWC Healthy Life-Style Check List (Part 1), Public Health Miyagi, vol., 341, pp.17-20, 2005. (in Japanese)
[3] M. Yamamoto, K. Mizuno, et al., An Approach of the Prevention Welfare Health Center in Tohoku Fukushi University, – Toward the Establishment of Self Health Management with SWC Healthy Life-Style Check List (Part 2), Public Health Miyagi, vol., 342, pp.17-20, 2005. (in Japanese)
[4] Sendai Wellness Consortium Research Report, The Project of Supporting of Creative Health Service, The Ministry of Economy, Trade and Industry, 2005. (in Japanese)
[5] Aoki, *A free clustering program*, http://aoki2.si.gunma-u.ac.jp/lecture/stats-by-excel/vba/html/clustan.html, (in Japanese).
[6] Lawrence D, Schank MJ. Health status, health perceptions, and health behaviors of young adult women. Int J Nurs Stud. 1993 Dec;30(6):527-35.
[7] Vaez M, Laflamme L. Health behaviors, self-rated health, and quality of life: a study among first-year Swedish university students. J Am Coll Health. 2003 Jan;51(4):156-62.
[8] Kim K, Uchiyama M, Liu X, Shibui K, Ohida T, Ogihara R, Okawa M. Somatic and psychological complaints and their correlates with insomnia in the Japanese general population. Psychosom Med. 2001 May-Jun;63(3):441-6.
[9] Ohida T, Kamal AM, Uchiyama M, Kim K, Takemura S, Sone T, Ishii T. The influence of lifestyle and health status factors on sleep loss among the Japanese general population. Sleep. 2001 May 1;24(3):333-8.