Food Recommendation System Using Clustering Analysis for Diabetic Patients

Maiyaporn Phanich, Phathrajarin Pholkul, and Suphakant Phimoltares*
Advanced Virtual and Intelligent Computing (AVIC) Research Center
Department of Mathematics, Faculty of Science, Chulalongkorn University
Pathumwan, Bangkok, Thailand
maiyaporn.p@student.chula.ac.th, phathrajarin.p@student.chula.ac.th, suphakant.p@chula.ac.th

Abstract— **Food and nutrition are a key to have good health. They are important for everyone to maintain a healthy diet especially for diabetic patients who have several limitations. Nutrition therapy is a major solution to prevent, manage and control diabetes by managing the nutrition based on the belief that food provides vital medicine and maintains a good health. Typically, diabetic patients need to avoid additional sugar and fat so the food pyramid is recommended to the patients for finding the substitution from the same food group. However, there is still a dietary diversity within food groups that can affect the diabetic patients. In this study, we proposed Food Recommendation System (FRS) by using food clustering analysis for diabetic patients. Our system will recommend the proper substituted foods in the context of nutrition and food characteristic. We used Self-Organizing Map (SOM) and K-mean clustering for food clustering analysis which is based on the similarity of eight significant nutrients for diabetic patient. At the end, the FRS was evaluated by nutritionists and it has performed very well and useful for nutrition area.**

Keywords- food recommendation system; clustering analysis; self-organizing map; k-mean clustering

I. INTRODUCTION

Nutrition therapy is a major solution to control diabetes by managing the nutrition based on the belief that food provides vital medicine and maintains a good health. Although medicine is our food, food is also our medicine. Some health problems require specific medication; many conditions of problems can be relieved effectively with nutrition therapy [1].

Not only a single disease, but also a group of autoimmune diseases is included in the diabetes. The human body usually needs sugar for energy; however, too much sugar in blood can vitally damage the body especially the diabetes. Therefore, the diabetes prevention would be the proper nutrition and healthy diet which balance sugar to the optimal level and maintain a healthy weight, respectively.

The Food Pyramid [2] is one of the choices recommended to the diabetic patients. Food items are classified by the nutrition. The same food group can substitute each other based on its nutrients. Nevertheless, the same food group from the food pyramid still has various details of significant nutrient that can affect to the diabetic patients. Moreover, human body requires multiple food groups more than the food pyramid groups. So the effective clustering from the various actual nutrients is needed to apply. The clustering will encourage diabetics to eat the widest possible variety of permitted food to ensure getting the full range of trace elements and other nutrients. Therefore, nutrition knowledge from the competent grouping adeptness has offered generally for nutrition therapy for diabetes.

In our recent survey, there are some related recommendations systems [3,4,5] in different purposes. For instance, Li et al. [6] proposed an ontology skeleton with hierarchical clustering algorithm (HCA) by using the class naming and instance ranking. This approach presents the new creating food groups and hierarchy of ranking food, however, this approach has manually set nutrients into three levels, that is, low, medium, high and the results of all food characteristics are considered. For this reason it is not efficient enough for grouping and providing the best result for diabetes diet care. Lee et al. [7] propose the personal diabetic food recommendation agent, including the ontology mechanism, the personal ontology filter and food fuzzy mechanism. This differs from our study that they presents only the remaining calories for dinner intake instead of all daily meals. Moreover, our approach emphasizes on substituted food yielding the recommendation system more efficient than the others. To improve the system better, the proposed system is not only for recommendation but also for demonstrating the substituted food for diabetic patients.

This paper is organized into seven sections. Section II introduces the theoretical foundation, which will provide the information about clustering techniques and diabetes used in this study. Section III describes the preparation of the data that will be used in the cluster analysis and the FRS afterwards. Section IV presents the detail of the food clustering by using SOM and K-means clustering. Section V shows the results received from the clustering analysis. Section VI illustrates the recommendation system and how to present the substituted food and clustering mechanism for diabetic diet care. Section VII presents case study of diabetic patients and simulates the situation using recommendation system. Section VIII discusses

*Corresponding author

978-1-4244-5943-8/10/$26.00 ©2010 IEEE

the performance of the system by invited nutritionists to complete the questionnaire and analyses the score from the results. At last, section IX draws the conclusions of FRS using food clustering analysis for diabetic patients.

II. THEORETICAL FOUNDATION

A. Clustering Analysis

Clustering analysis [8] is a common technique for statistical data analysis that used in various fields including machine learning, pattern recognition, and data mining. Clustering is a method of unsupervised learning which groups similar objects on the basis of their attributes into a same group called cluster. The purpose of clustering is to group the objects based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. In other words, the greater the similarity within a group and the greater the difference between the groups, the better the clustering. Nowadays, there are many clustering algorithms which can be classified into the following categories; Partitioning methods, Hierarchical methods, Density-based methods, Grid-based methods, and Model-based methods. The techniques used in this study are the competitive learning called Self-Organizing Map and K-mean clustering which are described in the following parts.

B. Self-Organizing Map

The Self-Organizing Map (SOM) [9] or commonly known as Kohonen network is a type of artificial neural network that is trained using unsupervised learning for the visualization and analysis of high-dimensional data purpose. It was invented by a Finnish professor named Teuvo Kohonen. The SOM composed of map units called nodes or neurons which are connected to adjacent neurons by a neighborhood relation. In the two-dimensional map, the neurons can be arranged in geometrical shape such as rectangle or hexagon as shown in Fig.1. The SOM algorithm will compute a model, m_i, for each node. These models are the representation of the input space of the training samples and organized into an n-dimensional ordered map in which similar models are closer to each other in the grid than the more dissimilar one. In this concept, the SOM is a similarity graph, and a clustering diagram, too.

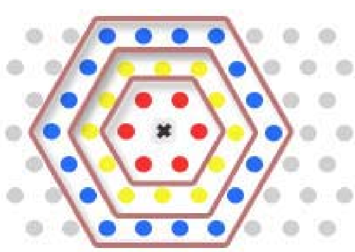

Figure 1. Neighborhood size 1, 2, and 3 of the centered node in a two-dimensional hexagonal grid.

The difference between the Self-Organizing Map and other artificial neural networks is that SOM use a neighborhood function to preserve the topological properties of the input space.

The SOM training can be considered to be a competitive learning since when the input vector is presented to the network, the Euclidean distance to all nodes in the map is calculated to find the node that gives the smallest distance is called best matching unit (BMU). Then, the weights of the BMU and its neighbor are adjusted towards the input vector. This adjustment process stretches the prototypes of the BMU and its topological neighbors towards the input vector. The BMU's local neighborhood can be determined by using the neighborhood radius which will shrink with time.

C. K-mean clustering

K-means clustering [10] is one of the most well known and commonly used partitioning clustering methods. The k-means algorithm takes the input parameter, k, and partitions a set of n objects into non-overlapping k clusters where $k<n$. This method aims to minimize the sum of squared distance between an object to the centroid which is called sum of squared error. The K-means algorithm proceeds as follows. First is to randomly select the k centroids from a dataset. The centroid represents the mean value of all the objects in the cluster. The next step is to assign the remaining object to the nearest cluster based on the distance between the object and the cluster mean, which is the centroid and then calculates the new mean for each cluster. This process iterates until there is no change of the centroids values. In other words, until the criterion function is convergence [11].

D. Diabetes and Diet

Diabetes Mellitus is a chronic disease that occurs when the body does not produce enough insulin or the cells ignore the insulin. The effect of uncontrolled diabetes is hyperglycemia or high blood glucose which can lead to many serious diabetes complications especially cardiovascular diseases but it can be prevented and controlled by managing the nutrition intake which is called nutrition therapy. Among the different components of nutrition, carbohydrates have the greatest influence on blood glucose levels so people with diabetes need to essentially concern the total amounts of carbohydrates consumed in each day [12]. For dietary purposes, carbohydrates can be classified as simple and complex carbohydrate. Simple carbohydrates or simple sugars are typically high in glycemic index which is the measure of the effects of carbohydrates on the blood glucose levels so they will cause a rapid rise in the blood glucose levels. Simple carbohydrates can be found in refined sugars and some fruits as well. Complex carbohydrates which are also called starch are usually found in grain product such as rice, noodles, and bread and starchy vegetables. Diabetic patients are able to consume complex carbohydrates food but with limitation and should avoid food with simple carbohydrates.

III. DATA PREPARATION

A. Food Dataset

This study is based on the dataset "Nutritive values for Thai food" provided by Nutrition Division, Department of Health, Ministry of Public Health (Thailand). The dataset gives eighteen nutrient values of food in 100 grams edible portion including Energy, Water, Protein, Fat, Carbohydrate, Fiber, Ash, Calcium, Phosphorus, Iron, Retinol, Beta-carotene, Vitamin A, vitamin E, Thiamine, Riboflavin, Niacin, and Vitamin C. The total number of data we used in this study is 290 and most of them are Thai local mixed food dishes and one plate dishes.

We have categorized the dataset into groups by two different ways.

1) Categorized by food characteristic: The dataset was divided into 22 groups based on their characteristic and shortly named A to V. The 22 groups consist of noodles, round rice noodles, Thai curry, northeastern food, namprik, rice, fried food, condiments, snack, fruits, Thai dessert, jam, milk and yogurt, juice, tea and coffee, soft drink, alcoholic drinks, chocolate/cocoa, vegetarian, nuts and beans, sausages, and miscellaneous. The first letter in the FoodID represents the group of that item e.g. A001 indicates that this item belongs to group A (noodles) and numbered 001. In Fig.2 shows the dataset.

2) Categorized by nutrition for diabetes: We have asked the nutritionist to help dividing the dataset based on the nutrients for diabetes. The dataset can be categorized into three groups as follows:

a) Normal Food (NF): Food in this group are very low in carbohydrate and have the least effect on the blood sugar level of the diabetic patients. Meat, poultry, meat fats, fish and seafood, vegetables excluding starchy vegetables, condiments with no sugar, and sugar-free herbal drinks are in this group.

b) Limited Food (LF): Foods in this group have large amount of carbohydrate but mostly are complex carbohydrates. Diabetic patients should carefully control the food intake in this group. Starchy food like rice, noodles, bread, and also starchy vegetables and grain are placed in this group.

c) Avoidable Food (AF): Foods in this group also have large amount of carbohydrate especially simple carbohydrate or simple sugar which is considered to primarily affect the blood sugar level. Diabetic patients should avoid food in the avoidable group e.g. sweets and chocolates, sweet fruits, and drinks that contain sugar.

According to two approaches of division, the nutritionist have been grouped the dataset. These two approaches are independent and different in objective. Because of the foods in the same group of food characteristic can be different in some nutrients so the food in the same food characteristic can be NF, LF, and AF. The subdivision is able to help the system work efficiently by suggesting the best food choices to the user. However, recommendation system will avoid presenting the food in Avoidable Food group which has a large amount of carbohydrates and recommend the food in Normal Food group and Limited Food group instead.

Food ID	Ngroup	English Name	Energy (Kcal)	Water (grams)	Protein (grams)	Fat (grams)	Carbohydrate (grams)	Fiber (grams)	Ash (grams)	Calcium (milligrams)	Phosphorus (milligrams)	Iron (milligrams)	Niacin (millgrams)	Vitamin C (millgrams)	Others
A001	2	Rice noodles with beef curry	131.00	75.00	6.60	7.40	9.40	0.30	1.60	59.00	63.00	0.40	1.60	0.00	...
A002	2	Rice noodles topped with chopped beef and gravy	112.00	76.20	3.50	4.10	15.20	0.40	1.00	41.00	42.00	0.20	1.30	1.00	...
A003	2	Rice noodles, fried, Thai style	239.00	53.80	7.70	12.30	24.30	0.60	1.90	150.00	109.00	2.00	1.70	1.00	...
A004	2	Rice noodles, fish with crab, with raw mungbean sprout	243.00	44.90	2.90	6.10	44.20	0.00	1.90	54.00	37.00	1.20	1.30	1.00	...
A005	2	Rice noodles with pork and liver	227.00	57.10	8.20	12.20	21.20	0.40	1.30	147.00	128.00	0.80	2.50	1.00	...
A006	2	Rice noodles with swamp cabbage, fish ball and sauce	72.00	84.20	2.80	2.70	9.20	0.20	1.10	98.00	18.00	0.50	0.70	1.00	...
A007	2	Rice noodles topped with shrimps, chinese kale, gravy	84.00	81.60	2.40	2.70	12.40	0.20	0.90	122.00	34.00	0.00	0.20	1.00	...
A008	2	Rice noodles topped with chicken, chinese kale, gravy	109.00	78.80	3.00	5.70	11.50	0.20	1.00	30.00	33.00	0.00	0.20	1.00	...
A009	2	Rice noodles topped with pork, chinese kale, gravy	113.00	77.70	2.90	5.40	13.00	0.10	0.90	0.00	31.00	0.00	0.10	1.00	...
A010	2	Rice noodles fried with pork, egg and soysauce	195.00	61.60	6.30	9.70	20.30	0.30	1.70	177.00	95.00	1.00	2.10	2.00	...
A011	2	Fine rice noodles with beefball	51.00	87.10	2.70	0.90	8.00	0.10	1.30	79.00	50.00	0.20	1.00	0.00	...
A012	2	Fine rice noodles fried with prok, egg, and soysauce	164.00	60.10	2.40	2.10	33.90	0.00	1.50	0.00	0.00	0.00	2.20	0.00	...
A013	2	Noodle, wheat, instant, with seasoning, prok flavoured	454.00	4.10	11.80	18.40	60.30	0.20	5.40	0.00	126.00	2.40	1.50	0.00	...
A014	2	Crispy fried fine rice noodles	505.00	9.10	2.90	29.40	57.10	0.20	1.50	6.00	47.00	0.60	2.00	2.00	...
A015	2	Fine rice noodles with coconut milk	173.00	63.30	3.90	6.60	24.50	0.50	1.70	12.00	64.00	0.70	0.90	1.00	...
A016	2	Fine rice noodles with coconut milk, and raw vegetable	140.00	70.70	2.80	5.50	19.60	0.00	1.30	20.00	40.00	0.70	2.60	0.00	...
A017	2	Fine rice noodles, fried, southern style	208.00	53.20	3.40	6.00	35.20	0.00	2.30	0.00	0.00	0.00	2.60	0.00	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
V004	1	Cashew nut seeds,baked	619.00	1.80	19.30	47.30	29.00	5.10	2.60	32.00	453.00	0.00	1.10	0.00	...

Figure 2. Nutritive values for Thai food in 100 grams edible portion dataset.

B. *Feature Extraction by Nutrient Ranking*

Nutrition plays an important part in the diabetes diet care and the importance of each nutrient is also different. For this matter, we have asked the nutritionist to rank the importance of eighteen nutrients to diabetic patients by giving a score ranging from 0-100 which 100 means the most important. The result is shown in Fig.3. The top eight nutrients are selected as main features to be included in the clustering analysis.

Nutrient	Energy	Water	Protein	Fat	Carbohydrate	Fiber
Score	95	25	80	85	100	90
Rank	2		5	4	1	3
Nutrient	Ash	Calcium	Phosphorous	Iron	Retinol	Beta-carotene
Score	30	60	45	50		
Rank						
Nutrient	Vitamin A	Vitamin E	Thiamine	Riboflavin	Niacin	Vitamin C
Score	55	75	65	40	35	70
Rank		6	8			7

Figure 3. Score and rank of the eighteen nutrients.

IV. FOOD CLUSTERING ANALYSIS

In this study, we used eight nutrients as attributes in the clustering according to the nutritionist recommendation. The eight nutrients include carbohydrate, energy, fat, protein, fiber, vitamin E, thiamine, and vitamin C. All of these nutrients have an influence on the diabetes diet in the different ways in both positive and negative. Also, these data have been normalized to the range of [0,1] before entering the training step.

Our cluster analysis consists of the two stages – first construct and train the SOM, and then the SOM is clustered using K-mean approach. The final result is the food clusters which foods in the same group provide the approximate amount of the eight nutrients.

A. *SOM Training*

The SOM learning algorithm is a double-loop algorithm consisting of two nested iterative loops. The inner loop is responsible for sequentially feeding the input into the network and adjusting the weights of the BMU and its neighborhood while the outer loop involves with decreasing the neighborhood radius and the learning rate.

In training the SOM, some parameters are required to set before the training begins: the number of neurons, dimension of the map grid, map shape, and neighborhood size. The number of neurons should usually be selected as large as possible, with the neighborhood size controlling the smoothness and generalization of the mapping [13].

In our study, we implemented the SOM using the Matlab SOM Toolbox [14] which set the default number of neurons to $5\sqrt{n}$ where n is the number of training data and the shape of our two-dimensional map grid is the hexagonal lattice.

The learning algorithm operates in the following steps.

1) Initialize each node weights randomly.
2) Select the input vector from the training data randomly and presented to the network.
3) Calculate the Euclidean distance between the input vector and every node to find the best matching unit (BMU).
4) Update the weight vectors of the BMU and its topological neighbor.
5) Repeat step 2 to 5 until all the training data is performed.
6) Increase t, denoted epoch, and repeat from step 2 until convergence occurs.

Before the training, the weight vectors of the same dimension of the training data for each neuron are initialized randomly. These random values are between 0 and 1.

The training step begins by fed the input vector to the network and traverse through each node in the map as in Fig. 4 and calculate the Euclidean distance using between weight vector and the current input vector to determine the BMU. The Euclidean distance is defined as follows:

$$Dist = \sqrt{\sum_{i=0}^{n}(v_i - w_i)^2} \tag{1}$$

where v_i is the input vector at dimension i and w_i is the weight vector at dimension i.

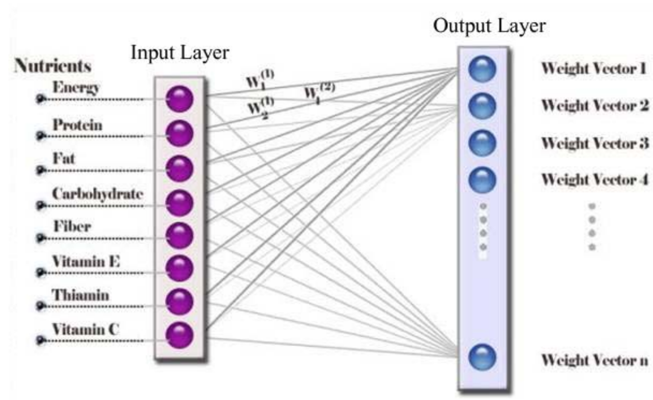

Figure 4. SOM networks where $W_i^{(j)}$ is weight vector j at dimension i and v = [Energy, Protein, Fat, Carbohydrate, Fiber, Vitamin E, Thiamine, Vitamin C].

The next step is to update the weights and the corresponding nodes in the neighborhood of BMU to make them closer to the input vector. Each node is updated by using the following equation (2)

$$W^{(j)}(t+1) = W^{(j)}(t) + h_{ci}(t)(s(t) - W^{(j)}(t)) \tag{2}$$

Where $W^{(j)}$ is the BMU's weight vector, $j = 1...k$ where k is the number of nodes, t is the epoch, s is the current input vector, and $h_{ci}(t)$ is the Gaussian function (3).

$$h_{ci}(t) = \alpha(t)exp\left(-\frac{dist^2}{2\sigma^2(t)}\right) \tag{3}$$

In (3), *dist* is a distance from a node to the BMU, and σ is the width of the neighborhood function which is decreasing monotonically over time, and $\alpha(t)$ is a learning rate where $0< \alpha(t) <1$

For this study, we use a power series function for the learning rate as defined in (4)

$$\alpha(t) = \alpha_0 \left(\frac{\alpha_T}{\alpha_0}\right)^{t/T} \quad (4)$$

where α_0 is the initial learning rate, α_T is the final learning rate, t is epoch and T the maximum number of epochs.

B. K-Means Clustering of the SOM

When the number of SOM units is large, to provide better quantitative analysis of the data, similar units need to be grouped [15]. Thus, the effective method, for this study is K-mean approach, is utilized to cluster the SOM for effective results.

The k-means algorithm consists of the following steps.

1) Randomly select k points as the initial centroids.
2) Assign each object to the nearest cluster.
3) Calculate the new centroid for each cluster.
4) Repeat step 2 and 3 until the centroids remain the same.

In step 2, the distances between each object to the centroids are calculated to determine the nearest cluster. Then, the mean for each cluster is recomputed to update the new centroid in step 3.

We clustered the SOM with different values of k, and use the Davies-Bouldin index [16] to evaluate the optimal k-value. The k-value is optimal when the related index is smallest

For this study, we used the values of k ranging from 15 to 20 and the index gives the smallest value for k =19.

Finally, a cluster number from 1 to 19 is assigned to each node in the map. Then, we mapped each training data to the closest node and assign the node's cluster number to that data. The final output of this clustering technique is a set of training data which implicitly defines clusters. The results of the clustering analysis are presented in the next section.

V. Results

The Self-Organizing Map is usually displayed in the form of unified distance matrix or u-matrix. The u-matrix [17] represents the SOM as a grid of neuron as illustrated in Fig.5. In Fig.5 is the u-matrix visualization of our maps which have dimension of 12x7. It also shows the distance between the adjacent neuron which is calculated and presented with different colors between adjacent nodes. A dark coloring between the neurons corresponds to a large distance and a lighter color indicates a smaller distance between them.

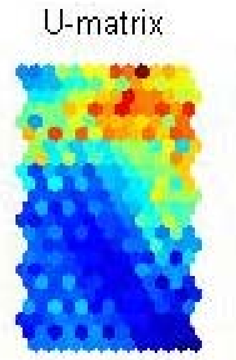

Figure 5. U-matrix visualization of the SOM.

In Fig.6, the visualization of the SOM of nutritive values for Thai food in 100 grams edible portion dataset. U-matrix is shown on top left, then component planes corresponding to the eight nutrients: Energy, Protein, Fat, Carbohydrate, Fiber, Vitamin E, Thiamine, and Vitamin C are sequentially illustrated. These nine subfigures are linked by position: in each figure, the hexagon in a certain position corresponds to the same map unit. In the U-matrix, additional hexagons exist between all pairs of neighboring map units. The color bar at right side of each subfigure indicates the value for each color represented in the node.

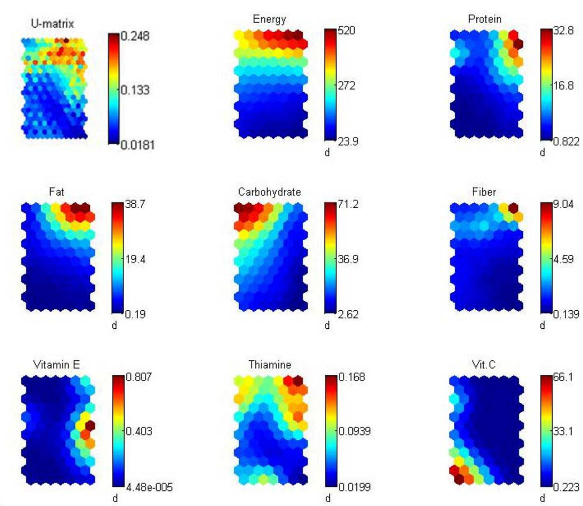

Figure 6. Visualization of the SOM of nutritive values for Thai food in 100 grams edible portion dataset.

Two evaluations used to measure the quality of the SOM are average quantization error and topographic error. In this study, the quantization error and the topographic error is equal to 0.1268 and 0.0103 respectively.

After the SOM is trained, we used K-means clustering to group similar units in the map together. Finally, the map is grouped into nineteen clusters based on the eight nutrients. The result of the clustering is illustrated in Fig.7.

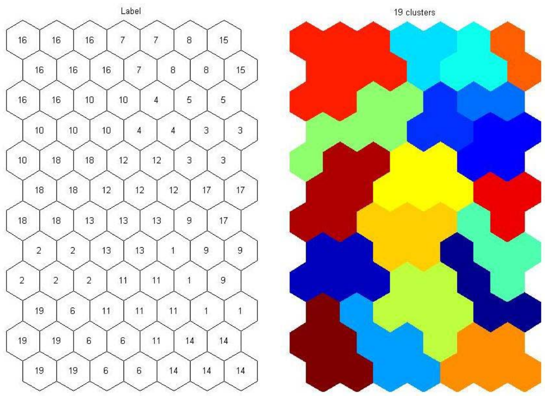

Figure 7. Visualization of the clustered map; (left) the cluster is assigned to a number between 1 and 19, (right) represent each cluster with different colors.

All of the 19 clusters show the unique properties for each group and can be defined in the meaningful way as described below in Table I.

TABLE I. DEFINITION OF EACH CLUSTER

Cluster	Definition		Food Group
	Low	High	
C1	Fiber, Thiamine	-	NF, LF, AF
C2	Protein, Vitamin E, Thiamine	-	LF, AF
C3	Vitamin C	Protein	NF, LF
C4	Vitamin E	Energy	NF, LF
C5	Vitamin E	Fat	NF
C6	Fat, Vitamin E	-	NF, LF, AF
C7	Vitamin C, Vitamin E	Energy, Carbohydrate	LF, AF
C8	-	Energy, Fat	NF, AF
C9	Carbohydrate	-	NF
C10	Vitamin E	Energy, Carbohydrate	LF, AF
C11	Thiamine	-	NF, LF, AF
C12	Vitamin C, Vitamin E	Carbohydrate	NF, LF, AF
C13	Vitamin C	-	NF, LF, AF
C14	Energy, Carbohydrate, Fiber, Thiamine	-	NF, LF, AF
C15	Vitamin C	Energy, Protein	NF, LF
C16	Vitamin E	Carbohydrate	NF, LF, AF
C17	Vitamin C	Vitamin E	NF, LF
C18	-	Carbohydrate	NF, LF, AF
C19	Fat, Vitamin E	-	NF, LF, AF

In addition, the distance matrix of the data within a group was calculated to determine the substituted food in the recommendation system. Finally, these results will be used as a database in the FRS as shown in the next section.

VI. FOOD RECOMMENDATION SYSTEM (FRS)

A. *Specification for Food Recommendation*

Recommended food (RF) for diabetic patients can be categorized into three groups as explained in section III:

1) *Normal food(NF):* The normal food is recommended for patients to eat normally.

2) *Limited food(LF):* This type of food is limited to consume. The patients are allowed to devour it a little bit per day.

3) *Avoidable food(AF):* The Avoidable food is a dangerous type of food. Users should avoid consuming it.

The FRS will recommend five sequential RFs to the patients by ranking from their proximal nutrients. When the users choose the NF, the FRS will recommend the substituted food only in the NF. Whereas when users select the LF and AF, the FRS will not suggest the AF but the food in other groups will be suggested. The result of the FRS will not advise the AF to the patients for ensuring that users will consume only the healthy food choices.

B. *Analyzing Nutrition Distance*

The FRS finds the recommended food by categorizing its nutrients. Furthermore, it aims to propose the RFs which not only have nearby nutrients, but also stay in the same food characteristic such as fruits, noodles, condiment. The FRS gathers the food characteristic from nutritionist. In the first step of recommendation system, users are required to select the food characteristic for efficiently searching the RF. After choosing the categories users would be able to select food name. The food name will be the key for retrieving the Distance Matrix (DM) of its clustered group in database.

In each DM, an element is a distance between a pair of food in the same cluster. This can rank nearby food items from the minimum distance. As a result, five foods are extracted and recommended ascendingly from ranking. It has been noted that the top of ranking is the food in the same category. If the number of recommended food is less than five, then the other recommended food are given by the food in different characteristic but still in the same cluster in ascending distance.

From Fig.8, food no.U002 is under consideration. In this example, all foods are in NF and the first alphabet of the name indicates the group of that food.

ID	G006	U002	U005	U006	V002	Z001
G006	0	0.273954	0.259877	0.233788	0.469015	0.54367
U002	0.273954	0	0.187653	0.1888	0.512995	0.6754
U005	0.259877	0.187653	0	0.189221	0.329492	0.4567
U006	0.233788	0.1888	0.189221	0	0.446334	0.5342
V002	0.469015	0.512995	0.329492	0.446334	0	0.9876
Z001	0.54367	0.6754	0.4567	0.5342	0.9876	0

Figure 8. Example of distances between proximate food in distance matrix.

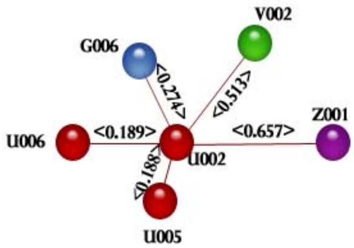

Figure 9. Distance between food no.U002 and five recommended food.

U002 have three foods in the same group, that is, U005, U006, and U002 itself. The foods in the other group are G006, V002, and Z001. The top ranking foods are the food from the same category as U002 so the foods must be U005 or U006. By the way, U005 has the distance less than U006 thus the first recommendation is U005 and the second is U006 while U002 itself is not included. When the foods from the same group are used up, the foods from the other groups, G006, V002 and Z001, are chosen with the minimum distances. The next minimum distance is 0.274 from G006, so G006 is the third recommended food. Consequently, V002 is the forth recommendation. Ultimately, Z001 is the fifth recommendation.

VII. Case Study of Diabetic Patients

With the reason that controlling diabetes especially to maintain a diet that suit for patient's condition is somewhat difficult. Diabetic patients, who have been suffered from diabetes, had been told by the doctor to be aware of their blood glucose level when the diagnosis shows that the level was higher than the limit and can cause more chance to develop a complication. It is also suggested that they have to be more careful in their diet but, usually, they cannot effective do as the suggestion because their daily food consumption behavior is barely changed. This problem can be solved by using the Food Recommendation System (FRS) that will help the patients to safely select the proper food and also provide more choices in consuming. With the simple user interface, the patients can use the FRS by only typing the name of the food that they usually take. In the case that the result received from the FRS indicates that this food is forbidden for diabetic patient, then recommend five other substituted foods yielding the approximate amount of eight nutrients and proper to eat. If they always use the FRS to select the food item as recommended, when they revisit the doctor to have a diagnosis, the test result came out to be very satisfied because the blood glucose level should be decreased to normal. Now, the patients themselves can select a variety of food items for each meal to maintain healthy diet every day.

VIII. Discussion

This system is evaluated by invited nutritionists to complete the questionnaire. The first part of the questionnaire is related to the format of presentation and the other part is involved with the acceptance of the FRS result. The nutritionists completed the questionnaires by assigning an integer value between 1~5 to each question as shown in Table II. '5' means being the most effectiveness whereas '1' suggests that the FRS need to have an amendment.

TABLE II. The Questionnaires and the Scaling

Questionnaire		
The Format of Presentation		**Score**
1	The Exquisite of Food Recommendation System	4.3
2	The Attractiveness for Using Food Recommendation System	3
3	Easy to Understand Usability of the System	3
4	The Amenity of the Result Format	3.3
5	The Clarity of Font, Size, Color and Other Complement	4.3
	Average Score	3.58
The Acceptance of the Result		
1	The Complacency in the Result	3.3
2	The Accuracy of the Result	4
3	The Perspicuity of the Result	3.3
4	The Rapidity to Execution	4.3
5	The Convenience for Using system	3.3
	Average Score	3.64

The Questionnaire Scaling : 5 - most effective; 4 - good; 3 - fairly good; 2 - fair; 1 - poor

The results from the questionnaires show that users are quite satisfied for using the FRS. The total score of the Presentation Format and the Acceptance of the Result are 3.58 and 3.64, respectively. For the Presentation Format score, the exquisite of the FRS and the clarity of complements are 4.3 defined as the almost effective system. It shows that the FRS is the user friendly system. Since the FRS provides a variety of results to the users, the FRS gets quite good score approximated 3.3 on the Amenity of the Result Format. However, users need to have some nutrition knowledge before using the FRS. Therefore, the FRS does not get high score on the Easy to Understand Usability. For the Acceptance of the Result, the Rapidity to Execution gains the highest score of 4.3. It shows that the FRS could rapidly response to the users. The accuracy of the FRS with score of 4 is also good for the users. The Complacency, the Perspicuity and the Convenience for using system receive 3.3 score above the fairly good score.

IX. Conclusion

For diabetes diet care, nutrition is the major key for control diabetes. Nevertheless, the existing categorization mechanism is not efficiently for classify the food group for diabetic patient. For the past research the ontology has been suggested for diabetes diet care and automated food mechanism, however, the results cannot achieve fairish food categories. For this reason, our research aims to present the next step in categorization by using the SOM algorithm along with K-mean clustering. In contrast with the existing research, SOM algorithm will categorize the food, by considering eight significant nutrients as main feature that have an effect on diabetic patients.

For the result of this research, a good recommendation for diabetes diet care is provided. This also develops the way for managing nutrition therapy knowledge.

For the future work, we expect to improve our research including with proposed service, technology and algorithms to diabetes diet care in Thailand.

REFERENCES

[1] J. P. Bantle, J. Wylie-Rosette, A. L. Albright, C. M. Apovian, N. G. Clark, M. J. Franz, B. J. Hoogwerf, A. H. Lichtenstein, E. Mayer-Davis, A. D. Mooradian, and M. L. Wheeler, "Nutrition recommendations and interventions for diabetes:a position statement of the american diabetes association," Diabetes care, vol. 31, pp. 61-78, January 2008.

[2] Food and Drug Administration's Food Pyramid. http://www.mypyramid.gov

[3] L. Huang, L. Dai, Y. Wei, and M. Huang, "A Personalized Recommendation System Based on Multi-Agent," Proc. 2nd Int'l Conf Genetic and Evolutionary Computing, pp. 223-226, September 2008.

[4] J. H. Kim, J. H. Lee, J. S. Park, Y. H. Lee, and K. W. Rim, "Design of Diet Recommendation System for Healthcare Service Based on User Information," Proc. 4th Int'l Conf Computer Sciences and Convergence Information Technology, pp. 516-518, November 2009.

[5] C. S. Lee, M. H. Wang, W. C. Sun, and Y. C. Chang, "Intelligent Healthcare Agent for Food Recommendation at Tainan City," Proc. IEEE Int'l Conf Systems, Man, and Cybernetics, pp.1465-1470, October 2008.

[6] H. C. Li and W. M. Ko, "Automated Food Ontology Construction Mechanism for Diabetes Diet Care," Proc. 6th Int'l Conf Machine Learning and Cybernetics, pp. 2953-2957, August 2007.

[7] C. S. Lee, M. H. Wang, H. C. Li, and W. H. Chen, "Intelligent Ontological Agent for Diabetic Food Recommendation," Proc. IEEE World Congress Computational Intelligence, pp. 1803-1810, June 2008.

[8] L. Chang-Shing , W. Mei-Hui, L. Huan-Chung and C. Wen-Hui, "Intelligent Ontological Agent for Diabetic Food Recommendation", IEEE Xplore., pp. 1803-1810, 2008

[9] J. Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques, Morgan Kaufmann, pp. 383-403, 2006.

[10] T. Kohonen, "The self-organizing maps," Proc. IEEE, vol. 30, no. 9, pp. 1464-1480, 1990.

[11] J. B. MacQueen : "Some methods for classification and analysis of multivariate observations," Proc. 5th Berkeley Symp. Mathematical Statistics and Probability, vol. 1, pp.281-297, 1967.

[12] K. Vipin, An Extensive Survey of Clustering Methods for Data Mining. http://www-users.cs.umn.edu/~han/dmclass/

[13] The Basic of a Healthy Diabetes Diet. http://diabetes.webmd.com/diabetes-diet-healthy-diet-basics

[14] J. Vesanto, SOM Implementation in SOM Toolbox. http://www.cis.hut.fi/projects/somtoolbox/documentation/somalg.shtml

[15] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-organizing map in Matlab: the SOM toolbox," Proc. Matlab DSP Conf., pp. 35-40, Nov. 1999.

[16] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," IEEE Trans. Neural Networks, vol. 11, no. 3, pp. 586-600, May 2000.

[17] D. L. Davies and D. W. Bouldin, "A cluster separation measure," IEEE Trans. Pattern Anal. Machine Intell., vol. 1, no. 4, pp. 224-227, 1979.

[18] J. Hollmen, Process Modeling Using the Self-Organizing Map, Master Thesis, Helsinki University , 1996.