On Map-Based Analysis of Item Relationships in Specific Health Examination Data for Subjects Possibly Having Diabetes

Naotake KAMIURA(1), Shoji KOBASHI, Manabu NII, Takayuki YUMOTO, Members, and Ichiro YAMAMOTO, Nonmember

SUMMARY In this paper, we present a method of analyzing relationships between items in specific health examination data, as one of the basic researches to address increases of lifestyle-related diseases. We use self-organizing maps, and pick up the data from the examination dataset according to the condition specified by some item values. We then focus on twelve items such as hemoglobin A1c (HbA1c), aspartate transaminase (AST), alanine transaminase (ALT), gamma-glutamyl transpeptidase (γ-GTP), and triglyceride (TG). We generate training data presented to a map by calculating the difference between item values associated with successive two years and normalizing the values of this calculation. We label neurons in the map on condition that one of the item values of training data is employed as a parameter. We finally examine the relationships between items by comparing results of labeling (clusters formed in the map) to each other. From experimental results, we separately reveal the relationships among HbA1c, AST, ALT, γ-GTP and TG in the unfavorable case of HbA1c value increasing and those in the favorable case of HbA1c value decreasing.

key words: specific health examination data, lifestyle-related diseases, self-organizing maps, hemoglobin A1c

1. Introduction

In Japan, the number of patients with diabetes has dramatically increased. In FY 2012, it was reported that the estimated sum of the number of persons strongly suspected of having diabetes and that of persons of which possibility having diabetes can not be denied was about 20.50 million [1]. The diabetes is one of the lifestyle-related diseases, and patients with diabetes are susceptible to various complications. Representative complications are diabetic neuropathy, diabetic retinopathy, and diabetic nephropathy. In addition, diabetes probably accelerates the progressions of cerebral apoplexy, cardiac infarction and ischemic heart disease caused by angiosclerotic [2]. The above complications degrade QOL (quality of life) of patients having diabetes, and the increase in the number of the patients leads to a marked increase in national medical care expenditure.

It has been recommended that, in Japan, middle-aged and aged national health insurance holders should take the specific health examination as measures against lifestyle-related diseases [3]. There are some items in specific examination data. The item hemoglobin A1c (HbA1c) is specially considered to be of importance when diagnosing diabetes. For subjects taking the specific health examination and living in Himeji city, Hyogo Prefecture, Japan, the percentage of the number of subjects of which the HbA1c values fall out of the appropriate range compared to the total number of subjects is 80 percent [4]. On the other hand, the percentage associated with Hyogo Prefecture is about 56 percent. Medical doctors that are members of Himeji Medical Association therefore try to analyze the situation occurring in Himeji city, and to take measures to overcome the problems associated with the above extremely high percentage.

In this paper, we explore relationships between HbA1c and other items in specific health examination data, considering it to be one of the basic researches establishing the above measures. The proposed method is based on self-organizing maps (SOM’s). SOM’s have strong research interest as a means of expressing and/or processing clinical examination results [5]–[9]. We construct maps available as useful tools visually classifying given data, by executing SOM learning. We focus on twelve items such as aspartate transaminase (AST), alanine transaminase (ALT), gamma-glutamyl transpeptidase (γ-GTP), and triglyceride (TG), in addition to HbA1c. We individually discuss the unfavorable case and the favorable case. In the former case, the HbA1c value deteriorates between successive two years. We then define the condition specified by body mass index (BMI), abdominal circumference (AC), and pick up the data from the examination dataset. In the latter, the HbA1c value measured in the preceding year is larger than that in the succeeding year, and we pick up the data according to the condition specified by BMI, AC and HbA1c. We calculate the difference between item values associated with successive two years, and generate training data by normalizing the values of the above calculation.

After SOM learning is complete, we label neurons in the map. We then employ one of the items of training data as a parameter. This labeling results in forming neuron clusters. We examine relationships between items by comparing such clusters to each other. From the experimental results, it is shown that the increase of TG tends to occur together with the increase of AST and ALT for the male subjects of HbA1c deteriorating, whereas we can divide female subjects into the group of subjects of which the HbA1c and TG values increase and the group of subjects of which the HbA1c, AST, ALT, and γ-GTP item values increase. On the other
hand, we clarify that, in the favorable case of HbA1c value decreasing, HbA1c has a similarity with each of AST, ALT, and γ-GTP. In addition, it is revealed that a part of patients of TG value decreasing overlap with a part of patients of which HbA1c item values are improved.

2. Preliminaries

Neurons are arranged in a map, and they have reference vectors. The number of elements in the data presented to the map is equal to the dimension of the reference vector. SOM learning is based on the following two formulas [10].

\[
\tau_i(t) = \tau_0(1-t/T),
\]

\[
W_i(t) \leftarrow W_i(t - 1) + \tau(t - 1)(X^l(t - 1) - W_i(t - 1)),
\]

where \(X^l(t)\) is the \(l\)-th training data, \(W_i(t)\) is the reference vector of the \(i\)-th neuron, \(\tau_i(t)\) is the learning rate at time \(t\), and \(T\) is the epoch number employed as the learning-termination condition. \(\tau_0\) in Eq. (1) and \(\tau(t)\) in Eq. (2) are an initial value and a product of the gauss function and Eq. (1), respectively. Let \(C_i\) denote the \(l\)-th neuron. If \(C_i\) is either a winner or one of the neurons considered to be neighbors around a winner, \(W_i(t)\) is modified according to Eqs. (1) and (2). We use Torus-SOM’s in this paper. SOM learning then modifies the reference vectors on condition that right and left edges and top and bottom edges of the map are continuous. Reference vectors in the edge region are well trained under the above continuity.

In this paper, we employ Torus-SOM referred to as My Tours. Its detailed usage of learning is given in [10]. After SOM learning is complete, neurons in a map are generally labeled. A method of neuron labeling is also given in [10]. For the map constructed by My Tours, we choose one of the elements in training data. The chosen element corresponds to one of the elements in the reference vector of the neuron, and labeling assigns a color to each of the neurons according to their values of the chosen element. Labeling then makes it possible to express the gradual variation between values of the chosen element for adjacent neurons with the difference in assigned colors. We recognize that some neuron clusters, which are specified by the types of color, are formed in the trained map after labeling. Note that labeling results depend on the choice of the elements. In other words, a result associated with neuron clusters under choosing some element changes when we choose another element for labeling. We can find the relationship among elements of the training data (i.e., neuron reference vectors), by comparing all of the labeling results with each other.

SOM’s thus make it possible to intuitively figure out potential attributes among data and to visually classify data. This seems to be suitable for informed consents, and motivates a number of medical doctors to apply SOM’s to various fields associated with analysis and classification of clinical examination results. Such SOM-based systems are proposed in [11]–[15]. The main developers of them are medical experts.

The elevation associated with glucose in circulating levels is one of the characteristics of diabetic patients. Since the item value of glucose frequently changes with every hour, HbA1c generally attract attentions of medical doctors when diagnosing some subject whether he/she has diabetes. The HbA1c value is considered to be the amount of glucose attached to hemoglobins in the red blood cells. The life-span of hemoglobin is estimated as about four months in average [16]. The hemoglobins run through blood vessels and move through the body, and blood sugars attach to them during the above span. We can predate the blood sampling date by one or two months and infer the situation of blood sugars, by measuring the HbA1c value.

Himeji Medical Association provided specific health examination data with the following items, systolic blood pressure (SBP), diastolic blood pressure (DBP), high density lipoprotein cholesterol (HDL cholesterol), low density lipoprotein cholesterol (LDL cholesterol), uric acid (UA) and creatinine in addition to HbA1c, BMI, AC, ALT, γ-GTP, and TG. In this paper, we choose some data, referring to item values associated with BMI and AC, and then generate training data presented to the map. Japan Society of Ningen Dock defines the health levels as shown in Table 1 [17]. The levels are specified by the item values. The level A means ‘no particular’. The level B corresponds to ‘slightly abnormal’. In the case where ‘follow-up observation and/or improvement of living custom’ are required, the level is considered to be C. The level D is the case of requir-

<table>
<thead>
<tr>
<th>Items</th>
<th>Health levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item values and health levels.</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>-18.5-24.9</td>
</tr>
<tr>
<td>AC (cm)</td>
<td>Male -84.9</td>
</tr>
<tr>
<td></td>
<td>Female -89.9</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
<td>-129</td>
</tr>
<tr>
<td></td>
<td>140-159</td>
</tr>
<tr>
<td>DBP (mmHg)</td>
<td>-84</td>
</tr>
<tr>
<td></td>
<td>90-99</td>
</tr>
<tr>
<td>AST (U/L)</td>
<td>0-30</td>
</tr>
<tr>
<td></td>
<td>36-50</td>
</tr>
<tr>
<td>ALT (U/L)</td>
<td>0-30</td>
</tr>
<tr>
<td></td>
<td>41-50</td>
</tr>
<tr>
<td>γ-GTP (U/L)</td>
<td>0-50</td>
</tr>
<tr>
<td></td>
<td>81-100</td>
</tr>
<tr>
<td>TG (mg/dL)</td>
<td>30-149</td>
</tr>
<tr>
<td></td>
<td>200-399</td>
</tr>
<tr>
<td>HDL cholesterol (mg/dL)</td>
<td>40-119</td>
</tr>
<tr>
<td></td>
<td>29-120</td>
</tr>
<tr>
<td>LDL cholesterol (mg/dL)</td>
<td>60-119</td>
</tr>
<tr>
<td></td>
<td>140-179</td>
</tr>
<tr>
<td></td>
<td>180-</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>-5.1</td>
</tr>
<tr>
<td></td>
<td>5.5-6.0</td>
</tr>
<tr>
<td>Creatinine (mg/dL)</td>
<td>Male -1.00</td>
</tr>
<tr>
<td></td>
<td>1.10-1.29</td>
</tr>
<tr>
<td></td>
<td>Female -0.70</td>
</tr>
<tr>
<td></td>
<td>0.80-0.99</td>
</tr>
<tr>
<td>UA (mg/dL)</td>
<td>2.1-7.0</td>
</tr>
<tr>
<td></td>
<td>7.6-8.9</td>
</tr>
</tbody>
</table>
ing ‘medical treatment and/or detailed examination’.

3. Map-Based Analysis for Items in Health Examination Data

In this section, we present a SOM-based method for analyzing the relationship between items in specific health examination data. It can be safely said that obesity probably increases the risk of being adult diseases such as arteriosclerosis, dyslipidemia, and hypertension. For a subject with too much visceral fat, the fat easily produces unesterified fatty acids. This situation increases blood concentration of fat, and tends to cause dyslipidemia. Besides, the physical reaction to insulin probably degrades for the subject. This results in increasing the probability of him/her having diabetes. We focus on subjects considered to be obese in addition to subjects with heavy visceral fat, and explore the relationship between HbA1c and other items in specific health examination data.

Japan Society of the Study of Obesity defines two cases where a subject is considered to be a patient on which weight reduction should be imposed as a medical treatment. The first case is for the subject with at least one of metabolic diseases related to obesity. On the other hand, the second case is for the subject with visceral fat accumulation, though it is diagnosed that he/she has no metabolic diseases related to obesity. In both cases, BMI value equal to or larger than 25.0 kg/m$^2$ is regarded as a condition for determining whether his/her weight must be reduced.

The abdominal CT is generally indispensable for precisely measuring the degree of visceral fat accumulation. There however exists a simple way to judge whether the subject has an unhealthy level of visceral fat. We simply judge the level, checking the value of abdominal circumference (AC). The degree of visceral fat accumulation is approximately proportional to the AC value. It is therefore considered that the visceral-fat level is unhealthy for male (or female) subject if his (or her) AC value is equal to or larger than 85 cm (or 90 cm) [18].

In the above context, we focus attention on BMI and AC values. We pick up the data out of the dataset provided from Himeji Medical Association according to the following condition: BMI $\geq 25.0$ and AC $\geq 85$ for the male subject (or AC $\geq 90$ for the female subject). In addition, for successive two years, we separately discuss the case of the HbA1c value increasing and that of the HbA1c value decreasing. Recall that, except for BMI and AC, each of the data consists of values of HbA1c, SBP, DBP, AST, ALT, γ-GTP, TG, HDL cholesterol, LDL cholesterol, UA and creatinine. The values of these items are measured from 2008 through 2012.

We next propose the generation of training data in the unfavorable case of HbA1c changing. Figure 1 illustrates the data generation for map learning. We first set $y$ ← 2008. In the first step, we choose male (or female) subjects on condition that BMI $\geq 25.0$ and AC $\geq 85$ (or AC $\geq 90$) hold for their data measured in $y$. In the second step, for each of the chosen subjects, we subtract an item value of his (or her) data measured in $y$ from the corresponding item value of his (or her) data in $y+1$. In other words, we calculate the difference between the data associated with successive two years, if the condition specified by BMI and AC item values is fulfilled for his (or her) data measured in $y$. Then, if the value measured in $y+1$ is smaller than the value measured in $y$ for some item, we set the subtraction result to 0 for that item. We apply the above to arbitrary items.

The difference calculation results in obtaining twelve-dimensional vectors. In the third step, we normalize element (i.e., item) values of these vectors according to the following formula.

$$x_j^l = (x_{\text{j-min}}^l)/(x_{\text{max}}^l - x_{\text{min}}^l),$$  \hfill (3)

where $x_{\text{j}}^l$ corresponds to the value of the $j$-th item in the $l$-th vector (i.e., the vector of the $l$-th subject) obtained by the difference calculation, $x_{\text{min}}^l$ (or $x_{\text{max}}^l$) means the minimum (or maximum) of all values associated with the $j$-th item (i.e., all $x_{\text{j}}^l$’s). $x_{\text{j}}^l$ is equal to or larger than 0. Recall that, if there is at least one subject having the $j$-th item value in the specific health examination measured in $y+1$ which is smaller than that measured in $y$, his/her $x_{\text{j}}^l$ takes the value of 0. Since the set of the data chosen according to the condition on BMI and AC includes the data of such subjects for any item, $x_{\text{min}}^l = 0$ holds. We then have the following formula.

$$x_j^l = x_{\text{j}}^l/x_{\text{max}}^l.$$  \hfill (4)

We employ $x_j^l$ as the $j$-th item value in the $l$-th training data. In the fourth step, we add the data normalized according to Eq. (4) to a training data set. The above are executed as long as $y$ is smaller than 2012.

![Fig. 1 Generation of training data set.](image-url)
4. Experimental Results

We train a map with fourteen rows and twenty columns, subject to the learning termination condition \( T = 100000 \). We set an initial value in Eq. (1) (i.e., \( \tau_0 \)) and an initial value of the neighborhood function represented by the gauss function to 0.01 and 30, respectively. Unless specifically mentioned, we use the above as training parameters. In this section, we separately discuss the relationships of HbA1c and other items for subjects of which the HbA1c values in \( y \) are smaller than those in \( y + 1 \), and for subjects of which the HbA1c values in \( y \) are larger than those in \( y + 1 \), where \( 2008 \leq y \leq 2011 \). The former case corresponds to the unfavorable case of HbA1c changing, while the latter is considered to be the favorable case. Note that to compare situations of items in the unfavorable case with those in the favorable case is beyond the scope of this paper.

4.1 Unfavorable Case of HbA1c Changing

We first show experimental results in the unfavorable case of HbA1c increasing. Himeji Medical Association provides 2557 data of male subjects and 1821 data of female subjects that fulfill the condition specified by BMI and AC values during the four-year period from 2008 through 2011. Recall that the condition is BMI \( \geq 25.0 \) and AC \( \geq 85 \) (or AC \( \geq 90 \)) for the male (or female) subject. Note that the condition is specified not only by BMI and AC item values but also by an HbA1c value. In the second step, we calculate the difference between the data measured in \( y \) and the data measured in \( y + 1 \) for each of the chosen subjects. We then consider the case to be favorable if the item value measured in \( y + 1 \) is smaller than that measured in \( y \). Therefore, when the subtraction result is a negative value for the \( j \)-th item in the \( l \)-th data, we set its absolute value to \( x^l_j \). On the other hand, when the subtraction results in a positive value, we give 0 to \( x^l_j \). Besides, for BMI, TG, HDL cholesterol, LDL cholesterol, and UA, there exit some exceptions for giving a positive value to \( x^l_j \). For these items, Table 1 shows that levels C and D appear not only in ranges where extremely large values belong to but also in ranges with comparatively small values. When the value of one of the five items in the \( l \)-th data measured in \( y + 1 \) belongs to the above small-value range, we set corresponding \( x^l_j \) to 0 regardless of the subtraction result. Note that \( x^l_j = 0 \) means \( x^l_j = 0 \) in the \( l \)-th training data.

Once the values of all possible \( x^l_j \)'s are determined, we normalize them in the third step, and add normalized vectors to a set to generate a training data set in the fourth step. Then, for any item of the data provided from Himeji Medical Association, since there is at least one case where \( x^l_j \) is equal to 0, we also use Eq. (4).

We construct a map according to the usage of learning given in [10], presenting members in the training data set generated by steps shown in Fig. 1. Figure 2 illustrates steps of obtaining labeling results and examining relationships between HbA1c and other items. We first set \( j \leftarrow 1 \). Note that the range of \( j \) is \( 1 \leq j \leq 12 \), because the number of items is twelve. Neurons in the map are next labeled, using \( j \)-th item as a parameter. Once twelve labeling results are obtained for the map, we compare neuron clusters formed under the condition that HbA1c is used as a parameter with those formed when other items are used. Let us assume that Fig. 3 depicts labeling results obtained according to two different parameters. We have a neuron cluster marked with \( \alpha \) (or \( \beta \)) in Fig. 3 (a) (or (b)). When comparing neuron clusters, we observe them in terms of their size, shape and position. If similarities are highly recognized between Clusters \( \alpha \) and \( \beta \), we can safely consider that there exists a strong relationship between the item used as the parameter for forming Cluster \( \alpha \) and that for forming Cluster \( \beta \). We thus clarify the correlation between HbA1c and other items.

Let us next discuss the favorable case in terms of HbA1c changing. Figure 1 also applies to this case. In the first step, we also pick up subjects with data fulfilling the following condition in \( y \), where \( 2008 \leq y \leq 2011 \): HbA1c \( \geq 5.6 \), BMI \( \geq 25.0 \) and AC \( \geq 85 \) (or AC \( \geq 90 \)) for the male (or female) subject. Note that the condition is specified not only by BMI and AC item values but also by an HbA1c value. In the second step, we calculate the difference between the data measured in \( y \) and the data measured in \( y + 1 \) for each of the chosen subjects. We then consider the case to be favorable if the item value measured in \( y + 1 \) is smaller than that measured in \( y \). Therefore, when the subtraction result is a negative value for the \( j \)-th item in the \( l \)-th data, we set its absolute value to \( x^l_j \). On the other hand, when the subtraction results in a positive value, we give 0 to \( x^l_j \). Besides, for BMI, TG, HDL cholesterol, LDL cholesterol, and UA, there exit some exceptions for giving a positive value to \( x^l_j \). For these items, Table 1 shows that levels C and D appear not only in ranges where extremely large values belong to but also in ranges with comparatively small values. When the value of one of the five items in the \( l \)-th data measured in \( y + 1 \) belongs to the above small-value range, we set corresponding \( x^l_j \) to 0 regardless of the subtraction result. Note that \( x^l_j = 0 \) means \( x^l_j = 0 \) in the \( l \)-th training data.

Once the values of all possible \( x^l_j \)'s are determined, we normalize them in the third step, and add normalized vectors to a set to generate a training data set in the fourth step. Then, for any item of the data provided from Himeji Medical Association, since there is at least one case where \( x^l_j \) is equal to 0, we also use Eq. (4).

We construct a map according to the usage of learning given in [10], presenting members in the training data set generated by steps shown in Fig. 1. Figure 2 illustrates steps of obtaining labeling results and examining relationships between HbA1c and other items. We first set \( j \leftarrow 1 \). Note that the range of \( j \) is \( 1 \leq j \leq 12 \), because the number of items is twelve. Neurons in the map are next labeled, using \( j \)-th item as a parameter. Once twelve labeling results are obtained for the map, we compare neuron clusters formed under the condition that HbA1c is used as a parameter with those formed when other items are used. Let us assume that Fig. 3 depicts labeling results obtained according to two different parameters. We have a neuron cluster marked with \( \alpha \) (or \( \beta \)) in Fig. 3 (a) (or (b)). When comparing neuron clusters, we observe them in terms of their size, shape and position. If similarities are highly recognized between Clusters \( \alpha \) and \( \beta \), we can safely consider that there exists a strong relationship between the item used as the parameter for forming Cluster \( \alpha \) and that for forming Cluster \( \beta \). We thus clarify the correlation between HbA1c and other items.

Let us next discuss the favorable case in terms of HbA1c changing. Figure 1 also applies to this case. In the first step, we also pick up subjects with data fulfilling the following condition in \( y \), where \( 2008 \leq y \leq 2011 \): HbA1c \( \geq 5.6 \), BMI \( \geq 25.0 \) and AC \( \geq 85 \) (or AC \( \geq 90 \)) for the male (or female) subject. Note that the condition is specified not only by BMI and AC item values but also by an HbA1c value. In the second step, we calculate the difference between the data measured in \( y \) and the data measured in \( y + 1 \) for each of the chosen subjects. We then consider the case to be favorable if the item value measured in \( y + 1 \) is smaller than that measured in \( y \). Therefore, when the subtraction result is a negative value for the \( j \)-th item in the \( l \)-th data, we set its absolute value to \( x^l_j \). On the other hand, when the subtraction results in a positive value, we give 0 to \( x^l_j \). Besides, for BMI, TG, HDL cholesterol, LDL cholesterol, and UA, there exit some exceptions for giving a positive value to \( x^l_j \). For these items, Table 1 shows that levels C and D appear not only in ranges where extremely large values belong to but also in ranges with comparatively small values. When the value of one of the five items in the \( l \)-th data measured in \( y + 1 \) belongs to the above small-value range, we set corresponding \( x^l_j \) to 0 regardless of the subtraction result. Note that \( x^l_j = 0 \) means \( x^l_j = 0 \) in the \( l \)-th training data.

Once the values of all possible \( x^l_j \)'s are determined, we normalize them in the third step, and add normalized vectors to a set to generate a training data set in the fourth step. Then, for any item of the data provided from Himeji Medical Association, since there is at least one case where \( x^l_j \) is equal to 0, we also use Eq. (4).

We construct a map according to the usage of learning given in [10], presenting members in the training data set generated by steps shown in Fig. 1. Figure 2 illustrates steps of obtaining labeling results and examining relationships between HbA1c and other items. We first set \( j \leftarrow 1 \). Note that the range of \( j \) is \( 1 \leq j \leq 12 \), because the number of items is twelve. Neurons in the map are next labeled, using \( j \)-th item as a parameter. Once twelve labeling results are obtained for the map, we compare neuron clusters formed under the condition that HbA1c is used as a parameter with those formed when other items are used. Let us assume that Fig. 3 depicts labeling results obtained according to two different parameters. We have a neuron cluster marked with \( \alpha \) (or \( \beta \)) in Fig. 3 (a) (or (b)). When comparing neuron clusters, we observe them in terms of their size, shape and position. If similarities are highly recognized between Clusters \( \alpha \) and \( \beta \), we can safely consider that there exists a strong relationship between the item used as the parameter for forming Cluster \( \alpha \) and that for forming Cluster \( \beta \). We thus clarify the correlation between HbA1c and other items.
learning is complete, we apply the neuron-labeling method, which depends on each of the twelve items, to each of the maps.

Figures 4(a), (b), (c), (d), and (e) depict neuron-labeling results obtained by using HbA1c, AST, ALT, γ-GTP, and TG as parameters, respectively. As mentioned in Sect. 2, we color neurons to express the degree of change associated with the value of the item for neuron labeling. The color ranges from blue to red as shown in Fig. 4(f). The lowest blue in Fig. 4(f) implies that we can recognize no change on the value of item used for labeling, and the change for the worse of the item during successive two years becomes larger as the color becomes redder. In this subsection, the color bar in Fig. 4(f) applies to labeling results.

Two clusters with pale blue appear in Fig. 4(a). The training data firing the neurons in the outside area of the clusters are generated from the specific health examination data of subjects whose HbA1c values measured in $y$ are approximately equal to or larger than those measured in $y + 1$, where $2008 \leq y \leq 2011$. On the other hand, for subjects with training data firing the neurons belonging to the clusters, we can recognize a certain amount of degradation in terms of HbA1c between $y$ and $y + 1$.

The cluster similar to the left cluster in Fig. 4(a) is formed in Fig. 4(b). In addition, there exist clusters similar to the left cluster in Fig. 4(a) in Figs. 4(c) and (e), while these clusters are quite faint compared with the cluster in Fig. 4(b). In Fig. 4(d), we can recognize no clusters similar to those in Fig. 4(a). This applies to the cases where we use BMI, SBP, DBP, UA, HDL cholesterol, LDL cholesterol, and creatinine as parameters for neuron labeling as shown in Fig. 5. It is thus revealed that, except for γ-GTP, HbA1c has relationships with items related to the liver such as AST and ALT. Besides, item values of HbA1c, AST and ALT simultaneously increase in $y + 1$ (i.e., the succeeding year among successive two years) for subjects of which the TG values deteriorate.

Figure 6 depicts some neuron-labeling results for female subjects. In Figs. 6(a), (b), (c), (d), and (e), neuron labeling is conducted on condition of using HbA1c, AST, ALT, γ-GTP, and TG as parameters for female subjects in case of HbA1c increasing. A slightly vague cluster also appears in its left part. Each of the clusters in Figs. 6(b)-(d) is similar to the right cluster in Fig. 6(a). On the other hand, we can recognize a similarity between the cluster in Fig. 6(e) and the left cluster in
Fig. 6 (a). It is thus revealed that female subjects tend to be divided into the following two group: the group with subjects of which the HbA1c and TG item values increase in the succeeding year and the group with subjects for which HbA1c, AST, ALT, and γ-GTP item values deteriorate.

For female subjects, we cannot recognize the relationships between HbA1c and each of BMI, SBP, DBP, HDL cholesterol, LDL cholesterol, UA, and creatinine. In other words, no clusters similar to Fig. 6 (a) appear as neuron-labeling results. Figure 7 depicts clustering results using SBP, DBP, UA, and creatinine as such examples. There exist two differences between male and female subjects. The first difference is related to γ-GTP. For female subjects, it is included in a set of items associated with the liver that increase in the succeeding year, together with HbA1c. The second difference is as follows: female subjects with values of TG deteriorating in the succeeding year differ from those with item values related to the liver deteriorating.

The subjects corresponding to data presented to the map are considered to be obesity. The obesity tends to cause fatty liver and/or diabetes. For persons with fatty livers, it is said that the values associated with ALT, AST, γ-GTP, and TG often deteriorate simultaneously. Recall that no cluster appears in Fig. 4 (d). Using SOM’s thus makes it possible to easily illustrate that the γ-GTP deterioration does not apply to the case of male subjects, who have both fatty livers and HbA1c values getting worse, picked up in this paper. In addition, neuron-labeling results in Fig. 6 visually clarify that, for female subjects picked up in this paper, there exist the group to which the subjects with HbA1c and TG values deteriorating belong and the group to which those with HbA1c and AST, ALT, and γ-GTP values deteriorating belong. To the best of our knowledge, no works have been reported on the above two groups of female subjects with HbA1c values increasing during successive two years.

We next explore the dependence on training parameters in forming neuron clusters. We label neurons in the maps constructed under training parameters changed from those used for the above-depicted maps. Figure 8 depicts some examples for female subjects. We set the map size to twelve rows and sixteen columns for the examples. Figures 8 (a)-(c) correspond to the case where changing parameters is just made for the map size. For the results in Figs. 8 (d)-(f), we change $\tau_0$ in Eq. (1) from 0.01 to 0.03. For those in Figs. 8 (g)-(i), we use 43 instead of 30 as an initial value of the neighborhood function. For those in Figs. 8 (j)-(l), the order of presenting training data is randomly changed from that for the map in Fig. 6. Figures 8 (a), (d), (g), and (j) depict results when using HbA1c as a parameter for neuron labeling. When applying AST-based labeling, we obtain results shown in Figs. 8 (b), (e), (h), and (k). The remaining results in Fig. 8 are obtained by using TG as a parameter for neuron labeling.
for neuron labeling. Recall that we employ Torus-SOM’s in this paper. A clear cluster and a slightly vague cluster appear in each of the HbA1c-based results. The positions of such clear clusters are similar to those of clusters formed in the case of using AST. We also recognize a similarity between each of such vague clusters and each of the clusters in Figs. 8 (c), (f), (i), and (l). The comparative relationship between labeling results in Fig. 6, thus, holds when changing training parameters in the above-mentioned manner.

### 4.2 Favorable Case of HbA1c Changing

In this subsection, we show experimental results in the favorable case of HbA1c decreasing. For specific health examination data, the total number of male subjects fulfilling the condition specified by HbA1c $\geq 5.6$, BMI $\geq 25.0$ and AC $\geq 85$ during the four-year period from 2008 through 2011 is 2495, while 1907 female subjects have the three item values fulfilling the following condition: HbA1c $\geq 5.6$, BMI $\geq 25.0$, and AC $\geq 90$. The number of training data is therefore 2495 for male subjects, and that of training data is 1907 for female subjects. Note that, when determining values of $x_{cl}^j$’s in Eq. (3) before the normalization, there exist exceptions for giving positive values to $x_{cl}^j$’s on BMI, TG, HDL cholesterol, LDL cholesterol, and UA as mentioned in Sect. 3.

Figure 9 depicts some results for male subjects. In Figs. 9 (a), (b), (c), (d), and (e), we label neurons, using HbA1c, AST, ALT, $\gamma$-GTP, and TG as parameters, respectively. We also assign colors to neurons in the results to express the degrees of change. In this subsection, though the color range is also from blue to red as shown in Fig. 9 (f), the change for the better of item values can become larger as the color becomes bluer. The lowest red in Fig. 9 (f) then implies that we can recognize no change on the value of item used for labeling. Note that the color bar in Fig. 9 (f) applies to labeling results.

Neuron labeling based on using HbA1c results in the clear cluster as shown in Fig. 9 (a). In Figs. 9 (b) and (e), the clusters are recognized in the position similar to that in Fig. 9 (a). In addition, faint clusters having a similarity with that in Fig. 9 (a) in terms of the position appear in Figs. 9 (c) and 9 (d). It is thus established that item values of AST, ALT, $\gamma$-GTP, and TG tend to simultaneously decrease in $\gamma + 1$ (i.e., the succeeding year among successive two years) for subjects of which the HbA1c value is improved. When using BMI, SBP, DBP, HDL cholesterol, LDL cholesterol, UA, and creatinine as parameters for neuron labeling, we can recognize no clusters each of which is similar to that shown in Fig. 9 (a). Figure 10 depicts some examples of such neuron labeling.

When we label neurons subject to using HbA1c, AST,
ALT, γ-GTP, and TG as parameters for female subjects, clusters appear in Figs. 11 (a), (b), (c), (d), and (e), respectively. We recognize a couple of clusters as shown in Figs. 11 (a)-(d). They are clearly similar to each other. It is thus clarified that, for female subjects, HbA1c substantially has relationships with items related to the liver. On the other hand, Fig. 11 (e) shows that two clusters seem to be formed as the labeling result. Although we cannot recognize the perfect resemblance between Fig. 11 (e) and each of the other results in Figs. 11 (a)-(d), a cluster close to the center in Fig. 11 (e) is similar to a right faint cluster in Fig. 11 (a). It is thus probable that HbA1c values are improved in the succeeding year among successive two years for a part of female subjects of which TG item values decrease.

Clusters similar to those in Fig. 11 (a) are not recognized when we conduct neuron labeling on condition that BMI, SBP, DBP, HDL cholesterol, LDL cholesterol, UA, and creatinine are used as parameters for female subjects. Some examples of labeling results obtained under such parameter selection are shown in Fig. 12.

5. Conclusions

In this paper, we presented a method of examining relationships between specific health examination data items, using self-organizing maps. We pick up HbA1c, BMI, SBP, DBP, AST, ALT, γ-GTP, TG, HDL cholesterol, LDL cholesterol, UA and creatinine as target items. We define the standard specified by BMI and AC in the case of HbA1c value increasing between successive two years, and choose the data out of a specific health examination dataset according to it. On the other hand, if HbA1c value decreases, we define the standard specified by HbA1c in addition to BMI and AC to choose the data. We then generate the training data by subtracting each item value associated with the preceding year among the two years from corresponding item value associated with the succeeding year, and by normalizing the subtraction results. We construct a map for male subjects and that for female subjects, and label neurons in each map. For neuron labeling, we employ item values as parameters. We observe neuron clusters formed in the map to analyze the relationships. We have shown that, in the unfavorable case of HbA1c deteriorating, the increase of TG tends to accompany increase of HbA1c, AST and ALT for the male subjects. It has also been established that female subjects can be split into two groups. One group consists of subjects of which the HbA1c and TG values increase. The female subjects of which the HbA1c, AST, ALT, and γ-GTP item values increase tend to belong to another group. In the favorable case of HbA1c being improved, we have shown that the change of HbA1c value is similar to that of each of the items related to the liver (i.e., AST, ALT, γ-GTP). In addition, it has been revealed that a part of patients with improved TG item values considerably overlaps with a part of patients with improved HbA1c item values.

In future studies, we will modify the proposed method so that it can be introduced into medical practice. In addition, we apply the proposed method to clinical examination results with more items than specific health examination data used in this paper, if such data are available.

Acknowledgments

This research was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (A) (Grant number JP25240038).

References

Naotake Kamiura received the B.E. degree (Electronic Engineering) in 1990, the M.E. degree (Electronic Engineering) in 1992 and the D.E. degree (Doctor of Engineering) in 1995 from Himeji Institute of Technology, Japan. He is currently a professor in the Department of Electronics and Computer Science, Graduate School of Engineering, University of Hyogo. His research interests include the application of soft computing to medical engineering. He is members of the Japanese Society of Medical Imaging Technology and the IEEE.

Syoji Kobashi received B.E. (1995), M.E. (1997), and Dr. of Engineering (2000), all from Himeji Institute of Technology, Japan. He is currently an associate professor at University of Hyogo since 2005, and a guest associate professor at Osaka University, WPI immunology frontier research center since 2011. He was a visiting scholar at University of Pennsylvania, USA, in 2011-2012. His research interests include medical image understanding. He published over 60 journal papers, 250 conference proceeding papers, and some invited papers. He received 15 awards including Franklin V. Taylor Memorial Award from IEEE-SMCS, 2009. He has been serving on the chair of International Forum on Multimedia and Image Processing in 2012, 2014 and 2016, Publication chair and organization chair of IEEE SMC2018, General co-chair of International conference on informatics, electronics & Vision (ICIEV), 2016, and others. He is the editor at large of autosoft journal, and an associate editor of 3 journals, and a guest editor of some special issues. He is a senior member of IEEE.

Manabu Nii received the B.E. degree in 1996, the M.E. degree in 1998 and the D.E. degree (Doctor of Engineering) in 2001 from Osaka Prefecture University, Japan. He is currently an assistant professor in the Department of Electronics and Computer Science, Graduate School of Engineering, University of Hyogo. He is members of Japan Society for Fuzzy Theory and Intelligent Informatics, the Institute of Systems, Control and Information Engineers and the IEEE.

Takayuki Yumoto received the BE, Master of Informatics, and PhD degrees in Informatics from Kyoto University, in 2002, 2004, and 2007, respectively. He is an assistant professor at University of Hyogo since 2007. His research interests include Web search and mining. He is a member of the ACM, the IEEE Computer Society, the Information Processing Society of Japan (IPSI), and the Database Society of Japan (DBSJ).

Ichiro Yamamoto received the M.D. and PhD degrees from Kawasaki Medical University in 1983 and from Graduate School of Medicine Dentistry and Pharmaceutical Sciences, Okayama University in 1987, respectively. He established Yamamoto Clinic of Internal and Gastroenterological Medicines in 1988, and has been engaged in community medicine for citizens of Himeji as the head-physician of his clinic. He is currently the president of Himeji Medical Association.