Clustering of Questionnaire Based on Features Extracted by Geometric Algebra

Minh Tuan Pham
Nagoya University
Furou-cho Chikusa Nagoya Aichi
464-8603 JAPAN

Kanta Tachibana
Nagoya University
Furou-cho Chikusa Nagoya Aichi
464-8603 JAPAN

Eckhard Hitzer
University of Fukui
3-9-1 Bunkyo Fukui-city Fukui
910-8507 JAPAN

Tomohiro Yoshikawa
Nagoya University
Furou-cho Chikusa Nagoya Aichi
464-8603 JAPAN

Takeshi Furuhashi
Nagoya University
Furou-cho Chikusa Nagoya Aichi
464-8603 JAPAN

***Abstract*—Clustering is one of the most useful methods to understand similarity among data. However, most conventional clustering methods do not pay sufficient attention to geometric properties of data. Geometric algebra (GA) is a generalization of complex numbers and of quaternions, and it is able to describe spatial objects and relations between them. In this study we introduce GA to systematically extract geometric features from data. We propose a new clustering method by using various geometric features extracted with GA. We apply the proposed method to clarification of human impressions of a product. In the field of marketing, companies often carry out a questionnaire on consumers for grasping their impressions. Analyzing consumers through the obtained evaluation data enables us to know the tendency of the market and to find problems and/or to make hypotheses that are useful for the development of products. Finally, we discuss clustering results of a questionnaire with/without GA.**

I. Introduction

Nowadays, it becomes more and more important for companies to investigate their customers' willingness to buy a product through a questionnaire. A typical questionnaire consists of various questions. Though it may include a direct question such as "Are you willing to buy it?", it is not clear whether the answer to the question is reliable. In this study, we propose a method to find latent willingness to buy on the assumption that if answering patterns to all questions are similar then the subjects have similar latent willingness.

We analyze a dataset of a questionnaire for a newly developed product. The characteristics of this dataset are:

1) The same m questions are asked for n different objects (usage scenes of the product);
2) Each subject answers according to his/her willingness to buy for either of three different prices, and does not answer for the other prices.

Considering the first characteristic, we regard a pattern of answering to the questions by a subject as a tuple of m points in a n dimensional space. This aims to extract features of n dimensional *shape* formed by the m vectors. For this purpose, we use geometric algebra (GA) which can describe spatial vectors and higher order subspace relations between them [1], [2], [3], to systematically undertake various kinds of feature extractions, and to predict latent willingness to buy of subjects. There are already many successful examples of its use in colored image processing or multi dimensional time-series signal processing with complex numbers or quaternions, which are low dimensional GAs [4], [5], [6], [7], [8], [9], [10]. And GA-valued neural network learning methods for learning input-output relationships [12] are well studied.

For the second characteristic, we utilize harmonic functions [11] for semi-supervised learning. In our proposed method, geometric features extracted with GA can be used for defining a weighted graph over unlabeled and labeled data where the weights are given in terms of a similarity function between subjects.

To evaluate the effect of features extracted with GA, we examine kernel matrices induced from the geometric features using kernel alignment [13] between them. This paper also reports a result of semi-supervised clustering of subjects taking geometric properties of a questionnaire into consideration.

II. Method

A. Feature extraction with GA

GA is also called Clifford algebra. An orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ can be chosen for a real vector space $\mathbb{R}^n$. The GA of $\mathbb{R}^n$, denoted by $\mathcal{G}_n$, is constructed by an associative and bilinear product of vectors, the geometric product, which is defined by

$$\mathbf{e}_i\mathbf{e}_j = \begin{cases} 1 & (i = j)\,, \\ -\mathbf{e}_j\mathbf{e}_i & (i \neq j)\,. \end{cases} \tag{1}$$

GAs are also defined for negative squares $\mathbf{e}_i^2 = -1$ of some or all basis vectors. Such GAs have many applications in computer graphics, robotics, virtual reality, etc [3]. But for our purposes definition (1) will be sufficient.

Now we consider two vectors $\{\mathbf{a}_l = \sum_i a_{li}\mathbf{e}_i, l = 1, 2\}$ in $\mathbb{R}^n$. Their geometric product is

$$\mathbf{a}_1\mathbf{a}_2 = \sum_{i=1}^{n} a_{1i}a_{2i} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(a_{1i}a_{2j} - a_{1j}a_{2i}\right)\mathbf{e}_{ij}, \qquad (2)$$

where the abbreviation $\mathbf{e}_{ij} = \mathbf{e}_i\mathbf{e}_j$ is adopted in the following. The first term is the commutative inner product $\mathbf{a}_1 \cdot \mathbf{a}_2 = \mathbf{a}_2 \cdot \mathbf{a}_1$. The second term is the anti-commutative outer product $\mathbf{a}_1 \wedge \mathbf{a}_2 = -\mathbf{a}_2 \wedge \mathbf{a}_1$. The second term is called a 2-blade. Linear combinations of 2-blades are called bivectors, expressed by $\sum_{I\in\mathcal{I}_2} w_I \mathbf{e}_I, w_I \in \mathbb{R}$, where $\mathcal{I}_2 = \{i_1 i_2 \mid 1 \le i_1 < i_2 \le n\}$ is the ordered combination set of two different elements from $\{1, \ldots, n\}$. For parallel vectors $\mathbf{a}_1 = \kappa\mathbf{a}_2, \kappa \in \mathbb{R}$, and the second term becomes 0.

Next, we consider the geometric product of $\mathbf{a}_1\mathbf{a}_2 = \mathbf{a}_2 \cdot \mathbf{a}_1 + \mathbf{a}_2 \wedge \mathbf{a}_1$ with a third vector $\mathbf{a}_3$. First, because $\mathbf{a}_1 \cdot \mathbf{a}_2 \in \mathbb{R}$, $(\mathbf{a}_1 \cdot \mathbf{a}_2)\,\mathbf{a}_3$ is a vector. Next, $(\mathbf{a}_1 \wedge \mathbf{a}_2)\,\mathbf{a}_3 = \left(\sum_{I\in\mathcal{I}_2} w_I \mathbf{e}_I\right)\sum_i a_{3i}\mathbf{e}_i$. For a certain $I = i_1 i_2$,

$$\mathbf{e}_I\mathbf{e}_i = \mathbf{e}_{i_1}\mathbf{e}_{i_2}\mathbf{e}_i = \begin{cases} \mathbf{e}_{i_1} & (i = i_2)\,, \\ -\mathbf{e}_{i_2} & (i = i_1)\,, \\ \mathbf{e}_{i_1 i_2 i} & (i_1 \ne i \ne i_2)\,. \end{cases} \qquad (3)$$

Therefore $(\mathbf{a}_1 \wedge \mathbf{a}_2)\,\mathbf{a}_3$ can be separated into a vector, i.e. the sum of terms corresponding to the first two lines of eq. (3), and a 3-blade $\mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \mathbf{a}_3$, i.e. the sum of terms corresponding to the bottom line of (3). A linear combination of 3-blades is called a trivector which can be represented by $\sum_{I\in\mathcal{I}_3} w_I \mathbf{e}_I$, where $\mathcal{I}_3 = \{i_1 i_2 i_3 \mid 1 \le i_1 < i_2 < i_3 \le n\}$ is the combinatorial set of three different elements from $\{1, \ldots, n\}$.

In the same way the geometric product $\mathbf{a}_1 \ldots \mathbf{a}_k, (k \le n)$ of linear independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ has as its maximum grade term the k-blade $\mathbf{a}_1 \wedge \ldots \wedge \mathbf{a}_k$. Linear combinations of k-blades are called k-vectors, represented by $\sum_{I\in\mathcal{I}_k} w_I \mathbf{e}_I$, where $\mathcal{I}_k = \{i_1 \ldots i_k \mid 1 \le i_1 < \ldots < i_k \le n\}$ is the combination set of k elements from $\{1, \ldots, n\}$. 1-blades are vectors with $\mathcal{I}_1 = \{i_1 \mid 1 \le i_1 \le n\}$. 0-blades are scalars with $\mathcal{I}_0 = \{\emptyset\}$. For the unit real scalar, many authors simply write $\mathbf{e}_\emptyset = 1$. For $\mathcal{G}_n$, $\bigwedge^k \mathbb{R}^n$ denotes the set of all k-blades and $\mathcal{G}_n^k$ denotes the set of k-vectors. The relationship between k-blades and k-vectors is $\bigwedge^k \mathbb{R}^n = \{A \in \mathcal{G}_n^k \mid \exists\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}, A = \mathbf{b}_1 \wedge \ldots \wedge \mathbf{b}_k\}$. The most general element of $\mathcal{G}_n$ is $A = \sum_{I\in\mathcal{I}} w_I \mathbf{e}_I$, where $\mathcal{I} = \bigcup_{k=0}^{n} \mathcal{I}_k = \mathcal{P}(\{1, \ldots, n\})$, and $\mathcal{P}(\cdot)$ denotes the power set. Concrete examples are the general elements of $\mathcal{G}_2$ which are linear combinations of $\{1, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_{12}\}$; and general elements of $\mathcal{G}_3$ which are linear combinations of $\{1, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_{12}, \mathbf{e}_{13}, \mathbf{e}_{23}, \mathbf{e}_{123}\}$. A general element of $\mathcal{G}_n$ can always be represented by $A = \sum_{k=0}^{n} \langle A\rangle_k$ with $\langle A\rangle_k = \sum_{I\in\mathcal{I}_k} w_I \mathbf{e}_I$, where $\langle\cdot\rangle_k$ indicates an operator which extracts the k-vector part. The operator that selects the scalar part is abbreviated as $\langle\cdot\rangle = \langle\cdot\rangle_0$.

The geometric product of $k\,(1 \le k \le n)$ vectors yields

$$\begin{aligned} \mathbf{a}_1 &\in \mathcal{G}_n^1 = \mathbb{R}^n = \bigwedge^1 \mathbb{R}^n, \\ \mathbf{a}_1\mathbf{a}_2 &\in \mathcal{G}_n^0 \oplus \bigwedge^2 \mathbb{R}^n, \\ \mathbf{a}_1\mathbf{a}_2\mathbf{a}_3 &\in \mathcal{G}_n^1 \oplus \bigwedge^3 \mathbb{R}^n, \\ &\vdots \\ \mathbf{a}_1 \ldots \mathbf{a}_k &\in \begin{cases} \mathcal{G}_n^1 \oplus \ldots \oplus \mathcal{G}_n^{k-2} \oplus \bigwedge^k \mathbb{R}^n & (\text{odd } k)\,, \\ \mathcal{G}_n^0 \oplus \ldots \oplus \mathcal{G}_n^{k-2} \oplus \bigwedge^k \mathbb{R}^n & (\text{even } k)\,. \end{cases} \end{aligned} \qquad (4)$$

Now we propose a systematic derivation of feature extractions from a set or a series of spatial vectors $\xi = \{\mathbf{p}_l \in \mathbb{R}^n, l = 1, \ldots, m\}$. Our method is to extract k-vectors of different grades k; which encode the variations of the features. Scalars which appeared in eq. (4) in the case of $k = 2$ can also be extracted.

For ξ, a set of n dimensional vectors, $n' + 1$ feature extractions are derived where $n' = \min\{n, m\}$. For $k = 1, \ldots, n'$,

$$\begin{aligned} \mathbf{f}_k(\xi) &= \{\langle \mathbf{p}_{l_1} \cdots \mathbf{p}_{l_k} \mathbf{e}_I^{-1}\rangle, I \in \mathcal{I}_k\} \in \mathbb{R}^{({}_mC_k)|\mathcal{I}_k|}, \quad (5) \\ \mathbf{f}_0(\xi) &= \{\langle \mathbf{p}_{l_1}\mathbf{p}_{l_2}\rangle\} \in \mathbb{R}^{({}_mC_2 + m)}, \quad (6) \end{aligned}$$

where, $|\mathcal{I}_k|$ equals to the number of combinations of k elements from n elements, and ${}_mC_k$ is the number of combinations when we choose k elements from m elements. $\mathbf{e}_I^{-1}$ is the inverse of $\mathbf{e}_I$. For $I = i_1 \ldots i_k$, $\mathbf{e}_I^{-1} = \mathbf{e}_{i_k} \ldots \mathbf{e}_{i_2}\mathbf{e}_{i_1}$. We denote by $\mathbf{f}_k$ a feature vector extracted by a feature extraction f_k. $\mathbf{f}_0$ is the scalar part in the geometric product of 2 vectors which was chosen from n vectors. $\mathbf{f}_k$ is the coefficient of k-blade in the geometric product of k vectors which was chosen from n vectors.

Below we apply the feature extractions with GA to the case of questionnaire data. Each subject gave evaluation values to the same 10 questions for 6 objects. We therefore regard a filled out questionnaire as $m(= 10)$ points in an $n(= 6)$ dimensional space: $\xi = \{\mathbf{p}_1, \ldots, \mathbf{p}_{10}\}, \mathbf{p}_l = \sum_i^6 x_{l,i}\mathbf{e}_i$ with $x_{l,i} \in \{-2, -1, 0, 1, 2\}$. Using GA, various kinds of feature extractions can be undertaken systematically. In this paper, we use 3 kinds of features which extracted from ξ with GA.

The simplest feature extraction f_1, which is coordinate value, also used in conventional methods, is

$$\begin{aligned} \mathbf{f}_1(\xi) &= [\langle \mathbf{p}_1\mathbf{e}_1^{-1}\rangle, \ldots, \langle \mathbf{p}_1\mathbf{e}_6^{-1}\rangle, \ldots, \\ &\qquad \langle \mathbf{p}_{10}\mathbf{e}_1^{-1}\rangle, \ldots, \langle \mathbf{p}_{10}\mathbf{e}_6^{-1}\rangle] \\ &= [x_{1,1}, x_{1,2}, \ldots, x_{10,5}, x_{10,6}] \in \mathbb{R}^{60}. \end{aligned} \qquad (7)$$

A second feature extraction f_0 uses inner product of 2 points,

$$\mathbf{f}_0(\xi) = [\langle \mathbf{p}_1\mathbf{p}_1\rangle, \langle \mathbf{p}_1\mathbf{p}_2\rangle, \ldots, \langle \mathbf{p}_{10}\mathbf{p}_{10}\rangle] \in \mathbb{R}^{55}. \qquad (8)$$

If 2 questions are correlated for 6 objects, then the corresponding element of $\mathbf{f}_0$ becomes large.

And a third feature extraction f_2 uses outer product of 2 points,

$$\mathbf{f}_2(\xi) = [\langle \mathbf{p}_1\mathbf{p}_2\mathbf{e}_{12}^{-1}\rangle, \ldots, \langle \mathbf{p}_1\mathbf{p}_2\mathbf{e}_{56}^{-1}\rangle, \ldots, \langle \mathbf{p}_9\mathbf{p}_{10}\mathbf{e}_{12}^{-1}\rangle, \ldots, \langle \mathbf{p}_9\mathbf{p}_{10}\mathbf{e}_{56}^{-1}\rangle] \in \mathbb{R}^{675}. \tag{9}$$

Each $|I_2|$ elements of $\mathbf{f_2}$ express independence of 2 questions and the direction of hyper-plane spanned by the 2 questions in the 6 dimensional space. If 2 questions are uncorrelated, then the corresponding element becomes large.

B. GA kernel and alignment

For feature extractions $f_k, k = 0, 1, 2$ with GA, we define a similarity funtion between two instances $i, j \in \{1, \ldots, p\}$ as

$$w_{ij;k} = \exp\left(-\frac{\|\mathbf{f}_k(\xi_i) - \mathbf{f}_k(\xi_j)\|^2}{\sigma_k^2}\right), \tag{10}$$

where, $\|\cdot\|$ is the Euclidean distance in the feature space. And a parameter σ_k is decided by Zhu et al [11] as described in section II-C. The kernel matrix $W_k = [w_{ij;k}]$ is a symmetric matrix with p rows and p columns.

In this study, we combine 3 kinds of feature extractions to cluster instances. The effect of combinating 2 feature extractions becomes small if their kernel matrices are aligned. The alignment [13] between 2 kernel matrices is defined as,

$$A(W_k, W_l) = \sum_{i,j} \widetilde{w_{ij;k}}\widetilde{w_{ij;l}} \in (0, 1] \tag{11}$$

where, $\widetilde{w_{ij;k}} = w_{ij;k}\left(\sum_{\imath,\jmath} w_{\imath\jmath;k}^2\right)^{-\frac{1}{2}}$ is the element of a matrix $\widetilde{W_k}$ which is normalized W_k so that its squared sum of all elements becomes 1.

In addition, cluster structure embedded in the data distribution is evaluated as alignment with identity matrix E,

$$A(W_k, E) = \frac{1}{\sqrt{p}} \sum_i \widetilde{w_{ii;k}} \in (0, 1]. \tag{12}$$

The alignment with E becomes 1 when similarity between any two different instances is 0, i.e. no cluster structure is embedded.

On the other hand, when binary label $y_i \in \{-1, 1\}$ is given for each instance, if instances with the same label are allocated near and those with the different label are allocated far, then the clustering result of this feature agrees with the labels. This is evaluated by the alignment between W_k and $Y = [Y_{ij} = y_i y_j]$,

$$\begin{aligned} A(W_k, Y) &= \frac{1}{p}\left(\sum_{i,j|y_i=y_j} \widetilde{w_{ij;k}} - \sum_{i,j|y_i\neq y_j} \widetilde{w_{ij;k}}\right) \\ &\in [-1, 1]. \end{aligned} \tag{13}$$

A good combination of two feature extractions has a low $A(W_k, W_l)$ value. And, if $A(W_k, E)$ is low and $A(W_k, Y)$ is high, f_k induces a feature space with rich cluster structure and agreement with the given labels.

C. Semi-Supervised Learning for Clustering

Semi-supervised learning is a problem to infer labels for unlabel data $U = \{l+1, \ldots, l+u = p\}$ or unknown unlabeled data when the label $y_i \in \{-1, 1\}$ of a part $L = \{1, \ldots, l\}$ of the instance set is known. In this paper, we solve a problem in the case of labeling U. The goal is to find a binary function $\gamma\colon U \to \{-1, 1\}$ so that similar points have the same label.

Referring Zhu et al [11], we decide a parameter σ of the kernel as the following. During making a minimum spanning tree over all data points with Kruskal's Algorithm, we label temporarily an unlabel data $a \in U$ to the same label as the labeled datum which connects with a for the first time. Then we find the median distance of tree edges that connect two instances with different labels. We regard this this distance d_o as a heuristic to the median distance between class regions. We arbitrarily set $\sigma = \frac{d_0}{3}$ following the 3σ rule of Normal distribution, so that the weight of this edge is close to 0.

Then, we construct a $p \times p$ symmetric weight matrix $W = [w_{ij}]$. The weight matrix can be separated as

$$W = \begin{bmatrix} W_{LL} & W_{LU} \\ W_{UL} & W_{UU} \end{bmatrix} \tag{14}$$

at the l-th row and the l-th column. For this purpose Zhu et al. proposed to first compute a real-valued function $g\colon U \to [0, 1]$ which minimizes the energy

$$E(g) = \sum_{i,j} w_{ij}\left(g(i) - g(j)\right)^2. \tag{15}$$

Restricting $g(i) = g_L(i) \equiv y_i$ for the labeled data, g for the unlabeled data can be calculated by

$$g_U = (D_{UU} - W_{UU})^{-1} W_{UL} g_L, \tag{16}$$

where $D_{UU} = \mathrm{diag}(d_i)$ is the diagonal matrix with entries $d_i = \sum_j w_{ij}$, for the unlabeled data. Then $\gamma(i)$ is decided using the class mass normalization proposed by Zhu et al.

Because either of three prices $\{\varphi_1, \varphi_2, \varphi_3 \mid \varphi_1 < \varphi_2 < \varphi_3\}$ is indicated to a subject when he/she indicates willingness to buy, we subdivide the subjects into three groups according to the indicated price. Then, we calculate γ^{φ_k} for each $k \in \{1, 2, 3\}$ regarding one group of subjects as labeled and the other groups as unlabeled. After that, we check the consistency of subject i, i.e. whether the willingness decreases weakly monotonously with the price.

$$\gamma^{\varphi_1}(i) \geq \gamma^{\varphi_2}(i) \geq \gamma^{\varphi_3}(i). \tag{17}$$

If subject i contradicts to this condition then we clear the label y_i and repeat the semi-supervised learning regarding such subjects as unlabeled from this time on. As shown in Fig. 1, we repeat this procedure until contradictions do not occur any more.

III. Experimental Result

A. Web Questionnaire Data

We used a web questionnaire data for a new product to extract features with GA and find the latent willingness to

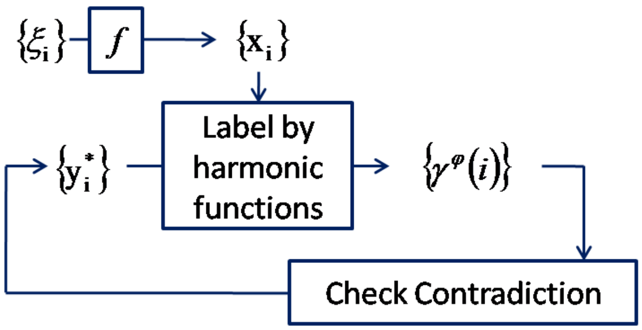

Fig. 1. Algorithm to find latent willingness to buy. From questionnaire data ξ_i, $f \in \{f_0, f_1, f_2\}$ extracts geometric features x_i. Label y_i^* is initially set to y_i. After calculating $\gamma^{\varphi}(i)$, labels of contradicted subjects are excluded from $\{y_i^*\}$. The algorithm ends when no more contradictions occur.

buy. Subjects answered 10 questions about each of 6 objects i.e. scenes in which the product was used. Subjects were asked to give an evaluation value to each question in 5 levels $\{1, 2, 3, 4, 5\}$, where "5" means "I agree very much" and "1" means "I disagree very much". Subjects answered willingness to buy the product for a price randomly selected from three prices.

We carried out the feature extractions with GA after subtracting 3 from all evaluation values so that $x_{l,i} = \{-2, -1, 0, 1, 2\}$. For simplicity the five level willingness values were binarized: "5", "4"$\mapsto 1$ and "3","2","1"$\mapsto -1$.

B. Kernel Alignment

We calculated alignment of kernel matrix W_k derived by feature $\mathbf{f}_k, k = 0, 1, 2$ which extracted with GA. Table I shows kernel alignmnent with other feature kernels, identity matrix E and the latent willingness matrix Y. The binary label $y_i = 1$ when the subject had latent willingness as a result of three analyses via different feature, and $y_i = -1$ otherwise, i.e. when the subject was judged not having latent willingness at least one analysis.

TABLE I
KERNEL ALIGNMENT EVALUATION

	W_0	W_1	W_2
W_0	1	0.92	0.63
W_1	0.92	1	0.51
W_2	0.63	0.51	1
E	0.71	0.82	0.36
Y	0.048	0.036	0.103

With kernel matrix W_1, kernel matrix W_2 had a smaller alignment (0.51) than W_0 had (0.92). Therefore, the effect of combination with feature extraction f_2 was better than with f_0. Also, the result showed that bacause both feature extractions f_0 and f_2 had lower alignment with E than f_1, they have more abundant structure of cluster between subjects. And they comtribute the total inference because alignment with Y is higher than f_1.

C. Web Questionnaire Analysis Results

TABLE II
RESULT WITHOUT GA

f_1	C	0.1%
	F-F	39.9%
	F-T	22.0%
	T-F	6.7%
	T-T	31.3%

Table II shows the result without introducing GA to find latent willingness to buy. In the table, "C" shows the percentage of subjects whose γ contradicted to condition (17) after the algorithm ended, thus we ignore those subjects. "F-F" shows the percentage of subjects whose $y_i = -1$, where, for simplicity, we do not mind what price was indicated to the subject, and $\gamma^{\varphi_1}(i) = -1$, i.e. the subject did not have either apparent or latent willingness even if the price was the lowest. "F-T" shows the percentage of subjects whose $y_i = -1$ but $\gamma^{\varphi_1}(i) = 1$, i.e. the subject answered not to have willingness but he/she had latent willingness at least for the lowest price. "T-F" shows the percentage of subjects whose $y_i = 1$ but $\gamma^{\varphi_1}(i) = -1$, i.e. the subject showed apparent willingness but from the similarity of answering patterns he/she was not willing to buy. "T-T" shows the percentage of subjects whose $y_i = 1$ and $\gamma^{\varphi_1}(i) = 1$, i.e. the subject had both apparent and latent willingness. The analysis made the following clear:

- Out of 38.0% of subjects ("T-F" or "T-T") who answered positively to the direct question of willingness, 31.3% of all subjects were detected as "truely" willing to buy ("T-T").
- Out of 61.9% of subjects ("F-F" or "F-T") who answered negatively to the direct willingness question, 22.0% of all subjects were detected as latently willing to buy ("F-T").

As a conclusion, 53.3% of subjects had a latent willingness to buy ("F-T" or "T-T"), and the other 46.7% subjects did not.

TABLE III
DETAILED LABELING RESULT

f_1(F-F) 39.9%		f_2	
		F-F	F-T
f_0	F-F	32.6%	0.2%
	F-T	7.0%	0.1%

f_1(F-T) 22.0%		f_2	
		F-F	F-T
f_0	F-F	6.0%	0.3%
	F-T	14.4%	1.3%

f_1(T-F) 6.7%		f_2	
		T-F	T-T
f_0	T-F	5.5%	0.1%
	T-T	0.9%	0.2%

f_1(T-T) 31.3%		f_2	
		T-F	T-T
f_0	T-F	3.0%	0.9%
	T-T	13.7%	13.7%

Next, we conducted a more detailed analysis introducing GA to define two more feature spaces which are based on f_0, f_2, respectively. Subjects were subdivided into the 5 groups of "C", "F-F", "F-T", "T-F" and "T-T" for each feature aspect. Thus Table III shows 4 sub-tables each of which further divides the corresponding subjects divided by f_1 to a matrix of judgements based on f_0 and f_2. Though 52.1% of subjects had the same result by all analyses based on f_0, f_1, f_2 (total of diagonal cells), the other 47.9% of subjects had different result. Especially,

- The second table shows that out of 22.0% of subjects who did not have apparent but had latent willingness according to analysis based on f_1 alone, only 1.3% of all subjects were judged as so according both to analyses based on f_0 and f_2.
- The bottom table shows that out of 31.3% of subjects who were judged as "truely" willing to buy in the analysis based on f_1 alone, 17.6% of all subjects were judged differently in at least one aspect of his/her answering pattern. On the other hand, the remaining 13.7% of all subjects can be judged as willing to buy with more confidence supported by the judgements based on f_0 and f_2.

As a conclusion, 15.0% of subjects were found to have latent willingness to buy ("F-T" or "T-T" by all analyses) with more confidence than in analysis without introducing GA.

Finally, we utilize principal component analysis (PCA) to visualize the data given by f_1. Figure 2 shows apparent and latent willingness. From figure 2, we can find clusters to which willing subjects belong.

IV. Summary

In this study, we introduced GA to extract geometric features from m-tuple of n dimensional vectors, and proposed 3 kinds of the kernels based on the extracted features. And we proposed the clustering algorithm by using the result based on the similarity funtion in each feature space. Then, we applied the proposed method to the clustering of answering patterns for a web questionnaire, deriving 3 kinds of feature extraction i.e. coordinates, outer products, and inner products. The result showed that feature extractions based on outer product and inner product, respectively, had more abundant structure of cluster between subjects and higher alignment with latent willingness to buy than in $m \times n$ dimensional vector space. Based on the extracted features, we found latent willingness to buy from the questionnaire data. The results showed that semi-supervised learning based on coordinates can detect subjects who have latent willingness to buy, and that introducing GA to the analysis can further find subjects who have *strong* latent willingness.

Acknowledgment

The authors thank to Dr. Sven Buchholz, University of Kiel. This work was supported by Grant-in-Aid for the 21st century COE program "Frontiers of Computational Science" Nagoya University and Grant-in-Aid for Young Scientists (B) #19700218.

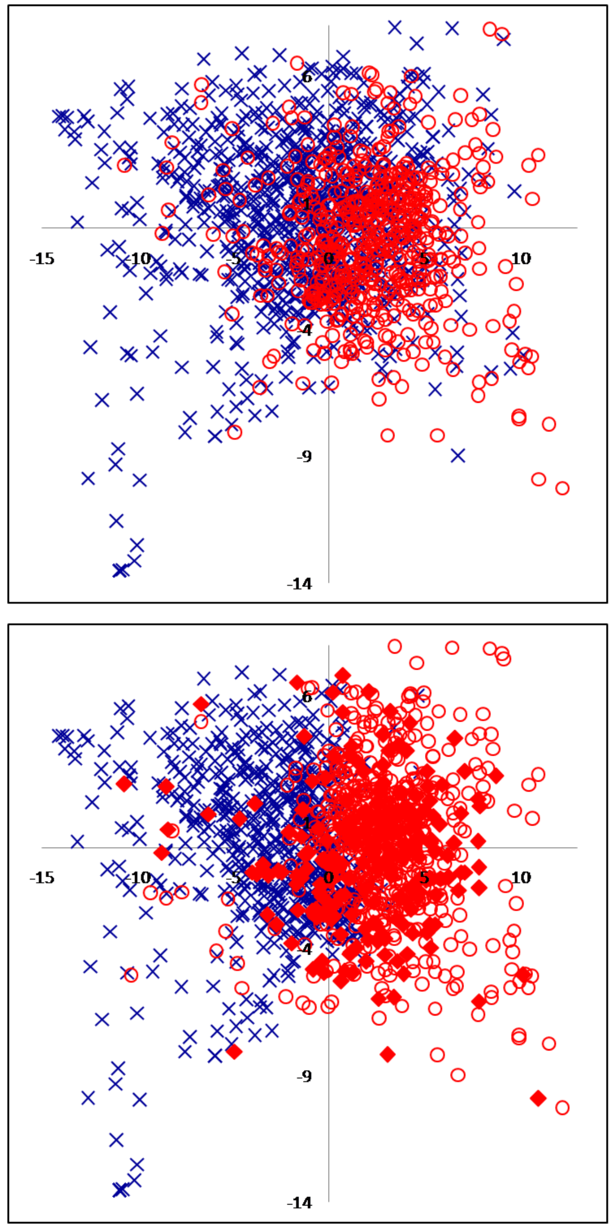

Fig. 2. The visualization of latent willingness to buy the product by PCA. The left shows apparent willingness and the right shows latent willingness to buy. A subject who did not have willingness is shown by '$\times$'. A willing subject is shown by '$\circ$'. The filled '$\diamond$' in the right figure show subjects who were judged as willing to buy with all feature extractions f_0, f_1, f_2.

References

[1] C. Doran and A. Lasenby, Geometric algebra for physicists, Cambridge University Press, 2003.
[2] D. Hestenes, New foundations for classical mechanics, Dordrecht, 1986.
[3] L. Dorst, D. Fontijne, and S. Mann, Geometric Algebra for Computer Science: An Object-oriented Approach to Geometry (Morgan Kaufmann Series in Computer Graphics), 2007.
[4] I. Sekita, T. Kurita, and N. Otsu, Complex Autoregressive Model for Shape Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 14, No. 4, 1992.

[5] A. Hirose, Complex-Valued Neural Networks: Theories and Applications, Series on Innovative Intelligence, Vol. 5, 2006.
[6] N. Matsui, T. Isokawa, H. Kusamichi, F. Peper, and H. Nishimura, Quaternion neural network with geometrical operators, Journal of Intelligent and Fuzzy Systems, Volume 15, Numbers 3–4, pp. 149–164, 2004.
[7] S. Buchholz and N. Le Bihan, Optimal separation of polarized signals by quaternionic neural networks, 14th European Signal Processing Conference, EUSIPCO 2006, September 4–8, Florence, Italy, 2006.
[8] T. Nitta, An Extension of the Back-Propagation Algorithm to Complex Numbers, Neural Networks, Volume 10, Number 8, pp. 1391–1415(25), November 1997.
[9] D. Hildenbrand and E. Hitzer, Analysis of point clouds using conformal geometric algebra, 3rd International conference on computer graphics theory and applications, Funchal, Madeira, Portugal, 2008.
[10] E. Hitzer, Quaternion Fourier Transform on Quaternion Fields and Generalizations, Advances in Applied Clifford Algebras, 17(3), pp. 497–517 (2007).
[11] X. Zhu, J. Lafferty, and Z. Ghahramani, Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. ICML 2003 workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining, 2003.
[12] G. Sommer, Geometric Computing with Clifford Algebras, Springer, 2001.
[13] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor, On kernel target alignment. Journal of Machine Learning Research, 2002.