Analyzing Kansei and design elements relations with PLS

Shigekazu Ishihara, Prof., Ph.D.[1], Mitsuo Nagamachi, Vice-Director, Prof., Ph.D.[2], and Keiko Ishihara, Assoc. Prof., Ph.D.[3]

1. Dept. of Kansei Design, College of Psychological Science,
Hiroshima International University
2. Yamaguchi University of Human Welfare and Culture
and User Science Institute, Kyushu University
3. Dept. of Communication, College of Psychological Science, HIU

i-shige@he.hirokoku-u.ac.jp

1. Introduction

Relations between Kansei word evaluation and design elements have been analyzed with Quantification Theory type I, which is created by Japanese statistician Chikio Hayashi in 1950s.

Evaluation values on a Kansei word are assigned to y (dependent) variable and design elements are assigned to x (independent) variables with dummy variables. In QT 1, the qualitative variable such as product body color is called "item". The variations of the item such like white / blue / red are called "categories". Categories are expressed with dummy variables those have 1/0 value. Weights to the categories will be found with a multiple regression model. Computing method is solving (number of all categories – 1) simultaneous equations (Hayashi, 1952).

QT 1 is deterministic because it is a variation of multiple regression model and its solving method uses least square method. Although QT1 is widely used, there are two defects: 1. Problem of sample size; QT1 incorporates (number of category – 1) dummy variables. In multiple regression model, simultaneous equations could not solved when number of variables exceed to the number of samples. Kansei engineering situation has tens of samples, and many cases have the larger number of design variables. Then, the analyst has to divide design variables to do analysis. 2. Problem of interactions between x variables; If there are heavy interactions between x

variables, analyzing result is distorted. The problem is known as a multicollinearity in multiple regression analysis.

PLS (Partial Least Squares) is becoming popular in chemometrics. PLS has possibility to resolve above problems. In this study we analyze personal garden Kansei engineering experiment data with PLS, and consider its analyzing ability.

In this paper, we present considerations on analysis ability of PLS with Kansei evaluation data of personal gardens.

2. Partial Least Squares

PLS was developed by Swedish econometrician Herman Wold and co-workers from middle of 1970s. Chemometrics is the PLS's most applied field from middle of 1990s. A typical example takes spectrum distribution on huge number of x. In these applications, number of x ups to several hundreds and correlations between x variables are very high because of spectrum. On the other hand, y takes measured value such as temperature or PH, and sample number is tens at the most. Brereton (2003) shows such smaller sample size cases in Chemometrics. Common multiple regression can not deal with such data.

PLS uses several latent variables. There are s (number of samples) observations of objective (dependent) variable. These become vector $\mathbf{y}$. There are p dimensional explanatory (independent) variables. These become vector $\mathbf{x}$. There are number s of $\mathbf{x}$, then become matrix $\mathbf{X}$. The algorithm given below is based on Miyashita and Sasaki (1995).

At the first step, $\mathbf{w}$, covariance vector of $\mathbf{y}$ and $\mathbf{x}$ is computed. The $\mathbf{w}$ is treated like eigenvector in principal component analysis. Second, latent variable t_1 is introduced. Output from t_1 ($t_1 = \Sigma x_{ik} w_k$, thus $\mathbf{t}_1 = \mathbf{X} \mathbf{w}_1$) is regarded as principal component score. Third, l_{11}, l_{12}, correlations between $\mathbf{x}$ and t_1 (these compose vector $\mathbf{l}_1$) are computed. They correspond to principal component loadings (correlation between principal component score and original variable). Forth, q_1, relation between $\mathbf{t}_1$ and $\mathbf{y}$ is computed. The q_1 is the result of single regression analysis (with no bias term), which takes $\mathbf{t}_1$ as an explanation variable and takes $\mathbf{y}$ as objective variable. Fifth, x-t_1-y relation can be computed. Sixth, second latent variable t_2 is introduced and we compute x-t_2-y relation with the same procedure noted above. This time, y takes the residual of x-t_1-y model, and $\mathbf{X}$ takes X residual of x-t_1-y model, which obtained by estimation by inverse way (Xnew = X - $t_1 l_1^T$). As the result, relations between two latent variables and y or x are obtained. Finally, we get a regression equation by composing these relations.
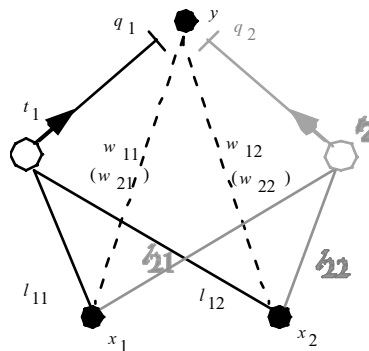

Figure 1. Structure of PLS

The high dimensional **x** is projected onto smaller dimension orthogonal space. The relation between the projection and *y* is solved with simple regression. Thus, the dimensionarity problem (sample size problem) is solved. The projection procedure is similar to procedure of principal component analysis. Since the projection is linear transformation, regression coefficients can be computed. Thus, correlations between explanation variables do not causes the multicollinearity problem. The multicollinearity is also avoided since there is no need for solving simultaneous equations.

3. Analysis of personal garden Kansei evaluation data and comparison with QT 1

  The mathematical features of PLS are quite attractive, but there is no statistical pointer for acceptance of the number of *x* variables. Since this study is the first attempt to use PLS in Kansei engineering, we should consider its ability.

  We have compared analyzing results of PLS and QT1, with Kansei evaluation data on personal gardens. We used 47 sample pictures of personal garden. Design element table has 32 items and 89 categories. PLS implementation which we used was JMP 5.2 (SAS).

  To analyze the data with QT 1, we have divided design elements into 5-fold (23/9/18/28/11 categories). We performed QT1 on each division, thus we got 5 results. We compared multiple correlation coefficients (correlation between predicted *y* values and measured values) of QT1 and PLS. Even when incorporating 89 variables, PLS's multiple correlation coefficient was much higher than QT1. Thus, PLS makes a model that fits the data better than QT1.
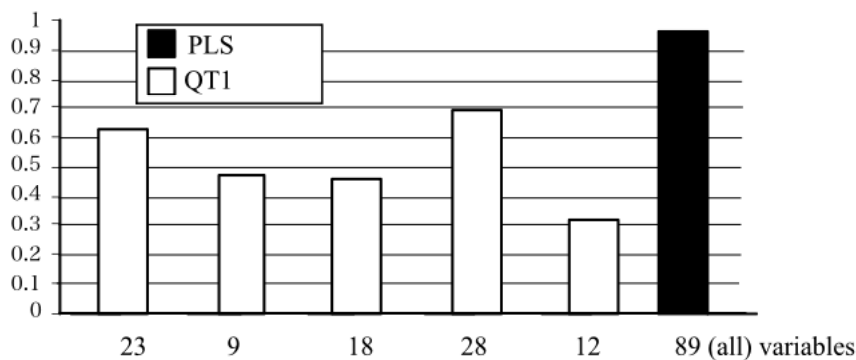

Figure 2. Comparison between PLS and QT1 results : Multiple Correlation Coefficients

  We got numerically excellent result with PLS. The analyzing result seem nearly perfect fit to the data. Another side of the perfect fit is the overfit to the data. Overfitting to the data is picking up all of the (unwanted) deviation of the sample and it makes the model too complex. As the result, overfitted model becomes more specific than the generalized.

  We also have compared QT 1 result with PLS result in qualitative manner. Although PLS can compute with all 89 *x* variables as we shown in above lines, in this comparison we compute the model by adding variable in 5 steps. Step 1 used 22 variables, step 2 used 22+9=31 variables, step 3 used 31+18=49, step 4 used

49+28=77, finally step 5 used 77+12=89 variables. The numbers of latent variables that used are four at step1 and 2, six at step 3, 4 and 5. These numbers were decided to have best performance (smaller residuals) by several trials with different number of latent variables.

Then we have surveyed accordance of ranks of categories in each item, between PLS and QT1. Figure 3 shows the accordance along to the number of variables.
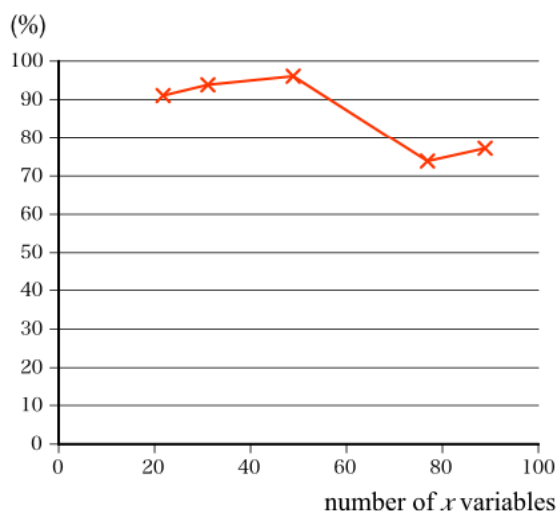

Figure 3. Accordance with PLS and QT1

When the number of x variables exceeds the sample size, the accordance slightly decreases (right side of the Fig.3). It seems that in the cases of smaller sample size, averaging effect is less, then smaller deviations reflecting to the result.

In many practical cases of KE, we can get numerically accurate model of relations between design elements and Kansei with PLS. In other word, analyzed result with PLS reflects smaller deviations (noise) by sample because of near perfect fit. PLS is very promising, but we have to read analyzing result carefully to decide the result reflects entire tendency or the particular sample.

References
Brereton, R., (2003), *Chemometrics*, Wiley, Chichester, England.
Hayashi, C., (1952), "On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view", *Annals of the Institute of Statistical Mathematics*, No.2, Vol.3, 69-98.
Miyashita, Y. and Sasaki, S., (1995), *Chemometrics-Chemical pattern analysis and multivariate analysis*, Kyouritsu Pressing, Tokyo (in Japanese).