The 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008)

Personalization Method of Multimedia Data Retrieval System Based on Impression

Teruhisa Hochin
Division of Information Science, Kyoto Institute of Technology
Goshokaidocho, Matsugasaki, Sakyo-ku, Kyoto-shi, Kyoto 606-8585, Japan
E-mail : hochin@kit.ac.jp

Abstract: This paper addresses to two major issues to the personalization of multimedia data retrieval system based on impression. One is on the decrease of the learning burden. The proposed method estimates the scores of an individual based on the grouping. The other is on the stability of the retrieval system. The proposed method uses a kind of rotation of a matrix. This method could keep the factors used in the system. This brings the stability to the retrieval system. By using these two methods, the user adaptation is easily, quickly, and stably attained.

1. Introduction

In recent years, retrieval of multimedia data has extensively been investigated[1-3]. Some research efforts treat the impression of multimedia data rather than their feature values and/or their meanings[1-3]. Here, impression means how we feel about a piece of multimedia data. For example, "fresh" and "clear" are the impression of a scenery picture. On the other hand, feature values are colors for pictures and pitch for sound clips. Meaning is, for example, "clear sky" for a picture. This paper treats the impression of multimedia data.

A method of accessing different kinds of media data directly and mutually has been proposed[3]. This method is called cross-media retrieval. In the cross-media retrieval, a kind of multimedia data is accessed based on another kind of multimedia data. For example, sound clips are directly accessed by specifying a picture, and vice versa. Users can freely navigate on the sea of multimedia data. The impression that users receive plays the important role in the system.

The straightforward way of learning user's favor is that a user evaluates all of the multimedia data stored in the system. This method has two major drawbacks. One is that the burden of the user is extremely high. No user can finish the evaluation because there are so many multimedia data in the system. The other drawback is that the system may not be able to keep the stability. The factor analysis is used in obtaining major factors describing impression. These major factors are called *Kansei factors*. Applying the factor analysis to the scores obtained through the learning does not always bring the same Kansei factors to the system. This means that the system may fall into the un-predictable state. From the view of software system, this state must be avoided. Another way is to use the learning phase repeatedly. In each learning phase, adaptation to the user gradually progresses. The drawback of this method is slow convergence. This process tends to be a time-consuming task. We need the method that takes short time to learn the user's favor for the purpose of avoiding the user from suffering heavy burden, and does not violate the stability of the system.

This paper proposes two methods. One is on the estimation of the scores of an individual for decreasing the user's learning burden. The multimedia data and the impression words carrying large amount of information are carefully chosen in order to be used in the learning. The scores of an individual are estimated only by answering a few questions using these multimedia data and the impression words. This could drastically decrease the learning burden of the user. The other is on the stability of the retrieval system. The proposed method uses a kind of matrix rotation. This can avoid applying the factor analysis again. This brings the stability to the retrieval system.

The remaining of this paper is as follows. Section 2 describes the derivation of Kansei factors. These are obtained through the factor analysis, and are described with a factor burden matrix. Section 3 proposes the estimation method of the individual's score matrix. Section 4 proposes the method estimating the factor burden matrix for an individual. Section 5 evaluates the proposed two methods. Finally, Section 6 concludes the paper.

2. Kansei Factors

The Semantic Differential (SD) method is the technique which uses subjective ratings of an idea, concept, or object by means of scaling opposite adjectives in order to study connotative meaning[4].

The scores form an (n, p) matrix Z, where n is the number of the target data, and p is the number of the impression word pairs. Applying the factor analysis to this matrix Z gives us a concise matrix. In the factor analysis, an (n, p) matrix Z is modeled by using the following equation:

$$Z = FA' + E \tag{1}$$

where F is an (n, m) matrix, A' is the transposed matrix of a (p, m) matrix A, and E is an (n, p) matrix. In obtaining the matrices F and A, m is selected to become as small as possible, and E is decided to become sufficiently small. The matrix A is called *the factor burden matrix*. The matrix F is called *the factor score matrix*. The matrix E is called *the error matrix*. The factors obtained are called *the Kansei factors*.

We have used forty scenery pictures and sixteen impression word pairs[3]. The impression word pairs are "bright - dark", "glad - sad", "warm - cold", "tense - relaxed", "large - small", "strong - weak", "heavy - light", "active - passive", "beautiful - ugly", "fresh - stale", "natural - artificial", "wet - dry", "bold - delicate", "clear - hazy", "simple - complex", and "sharp - dull". The five Kansei factors are obtained.

Therefore, in our case, $n = 40, p = 16$, and $m = 5$. The five factors are brightness, potency, activity, naturalness, and sharpness.

3. Estimation of Score Matrix

The first issue that we address to is the estimation of the individual's scores to the impression words for multimedia data. We propose a method based on the grouping. First, the scores are categorized into several major groups. In each major group, the subjects are again categorized into a subject group. For each major group, the representative picture and the impression word used in the learning are chosen. The score put to the impression word for the picture decides the subject group that the user belongs to. All of the scores are estimated by evaluating the pictures of all of the major groups.

3.1 Grouping pictures and subjects

The cluster analysis is applied to the forty pictures used in the cross-media retrieval system[3] to categorize them into several groups. The analysis is based on the Ward's method, and uses the Euclid square distance as the dissimilarity. Five groups are obtained. These groups are referred to as Picture Group A to E. The numbers of the pictures of these groups are 6, 6, 11, 10, and 7, respectively.

Next, the subjects, the number of which is thirteen, are categorized into several groups according to the similarity of the sensitivity for each picture group by using the cluster analysis. Here, we let the number of groups be equal to or less than five, which is the levels: 2, 1, 0, -1, and -2. Four groups are obtained for Picture Group D. Three groups are obtained for the other picture groups.

Each picture group has three or four subject groups. There are three to five subjects in each subject group. The scores to an impression word pair are not the same one for all of the subjects in a subject group. We have to decide the representative value of the score. The mode, which is the most frequent value in a collection of values, is adopted as the representative value of a score. If there are two or more modes, their average value is used as the representative value.

3.2 Selection of pictures and impression word pairs

There are $16*n$ parameters for n pictures in a picture group because there are scores for sixteen word pairs to a picture. We find the most descriptive parameter from $16*n$ ones. The most descriptive parameter is the parameter whose scores are different from one another for subject groups. The followings are the criteria choosing the most descriptive parameter.

(a) The parameter, whose representative values are different from one another for all of the picture groups, is selected.

(b) The parameter does not include the representative values of the other picture groups as possible.

One parameter could be selected for Picture Group B, C, and D, while two parameters are required for Picture Group A, and E. Table 2 shows the pictures and the impression word pairs (parameters). Divisions of scores according to subject groups are also shown in Table 2. The scores in the braces { and } are in the same subject group.

Table 1. Representative pictures, impression word pairs, and divisions of scores for picture groups

Group	Picture	Word pair	Score groups
A	Japanese garden	tense - relaxed	{2,1,0}, -1, -2
	Mountain and river	large - small	2, 1, {0,-1,-2}
B	Blue flower	warm - cold	{2,1}, {0,-1}, -2
C	Small flowers	simple - complex	2, {1,0}, {-1,-2}
D	Decorated tree	tense - relaxed	2, 1, {0,-1}, -2
E	Blue sky and houses	simple - complex	{2,1,0}, -1, -2
		active - passive	2, {1,0}, {-1,-2}

3.3 Deciding subject group

A user puts a score to a parameter, that is, to an impression word pair for a picture. The score decides the subject group that the user belongs to. For example, when a user puts 1 to the impression word pair "warm - cold" for the picture "Blue flower," the subject group of this user is decided to be the one having the scores 2 and 1 for the picture group B as shown in Table 2.

For the picture groups having two parameters, i.e., Picture group A and E, the decision procedure of the subject group is as follows:

(a) Select the most frequent score in the subject group. Calculate the value obtained by subtracting the ratio of the score in the other groups from the ratio of the score in the subject group.

(b) Following the values calculated in (a), the order of priority of the score is decided. It is decided that when the score is selected, the subject group corresponding to the score is chosen in the order of priority.

(c) If there are some combination of scores that can not decide the subject group, the subject group having the least distance between the scores and the scores selected in (a) is selected for each combination.

When the subject group is decided, the scores of the subject group in the picture group are set to the score matrix of the user. After all of the subject groups are decided, all of the scores of the score matrix is decided.

4. Estimation of Factor Burden Matrix

The next is a method of estimating a factor burden matrix.

The factor analysis is based on Eq. (1). When the matrix of average scores Z_{ave} is used, matrixes F_{ave}, A'_{ave}, and E_{ave} satisfying Eq. (2) are obtained.

$$Z_{ave} = F_{ave}A'_{ave} + E_{ave} \tag{2}$$

Similarly, when a matrix of an individual's scores Z_i is used, matrixes F_i, A'_i, and E_i satisfying Eq. (3) are obtained.

$$Z_i = F_iA'_i + E_i \tag{3}$$

Let us assume two (m, m) matrixes: M_i and N_i, where M_i satisfies $F_i = F_{ave}M_i$, and N_i satisfies $A'_i = N_i A'_{ave}$. Eq. (3) can be described with Eq. (4) by using M_i and N_i.

$$Z_i = F_{ave}M_iN_iA'_{ave} + E_i \tag{4}$$

When M_iN_i is represented with P_i, Eq. (4) becomes Eq. (5).

$$Z_i = F_{ave}P_iA'_{ave} + E_i \tag{5}$$

P_i can be obtained by omitting E_i, whose elements are small, from Eq. (5).

$$P_i = F^{+}_{ave}Z_iA'^{+}_{ave} \tag{6}$$

Here, F^{+}_{ave} (A'^{+}_{ave}, respectively) is the pseudo-inverse matrix of F_{ave} (A'_{ave}). An (n, m) pseudo-inverse matrix X of an (m, n) matrix A is the generalized inverse matrix satisfying $AXA = A, XAX = X, (AX)^* = AX$, and $(XA)^* = XA$, while a generalized inverse matrix satisfies only $AXA = A$.

As F^{+}_{ave} and A'^{+}_{ave} can be calculated from F_{ave} and A'_{ave} in advance, we can obtain P_i when Z_i is given. We can follow the individual's taste without applying the factor analysis again. This results in the stability of the system.

When P_i is a unit matrix, Z_i becomes Z_{ave}. The matrix P_i is considered to represent the individual's taste. We call P_i a *taste matrix*. Consider Eq. (5) as $Z_i = F_{ave}(P_iA'_{ave}) + E_i$. This means that the factor burden matrix for an individual can be obtained by multiplying the matrix P_i by the matrix A'_{ave}, which is the factor burden matrix obtained from the average score matrix. In this case, factors are considered to be rotated for the individual with keeping the factor score matrix F_{ave} as it is.

An Example of a taste matrix P_i is as follows.

$$P_i = \begin{pmatrix} 0.75 & 0.16 & 0.09 & -0.02 & -0.16 \\ 0.11 & 0.42 & 0.06 & -0.17 & -0.13 \\ 0.07 & 0.06 & 0.59 & -0.06 & -0.12 \\ 0.08 & -0.12 & -0.06 & 0.62 & -0.01 \\ -0.04 & 0.01 & 0.07 & 0.08 & 0.81 \end{pmatrix}$$

This matrix is calculated from the score matrix obtained by evaluating pictures. Absolute values of the diagonal elements of P_i are relatively large, while those of the others are small.

5. Evaluation

5.1 Evaluation methods

(1) Evaluation method of score matrix estimation

First, the amount of the learning burden is evaluated. The times of the proposed learning and the full learning are measured for a subject. The full learning means that the subject has to evaluate all of the impression word pairs for all of the pictures. In our case, the subject has to put 640 scores because there are sixteen impression word pairs for forty pictures.

The fitness of the retrieval to an individual is also evaluated. After five subjects finished the learning procedure, the subjects evaluated the fitness of the retrieval by comparing the results of the retrieval using the individual scores estimated through the learning with those using the average scores. The subjects used seven levels of scores: 3 to -3. The score 3 means that the retrieval extremely fits with the individual. The score -3 means that the retrieval does not fit with the individual at all. The score 0 means that the fitness of the retrieval is the same as that of the retrieval using the average scores.

(2) Evaluation method of factor burden matrix estimation

The proposed method of estimating a factor burden matrix is evaluated by retrieving pictures through the impression words. The impression words are ”bright,” ”large,” ”active,” ”beautiful,” and ”sharp.” These are the representative ones of the Kansei factors for pictures: brightness, potency, activity, naturalness, and sharpness, respectively[3]. The forty pictures, which are used in the retrieval system based on impression[3], are used. A subject selects the pictures corresponding to the impression words mentioned above in advance. These pictures constitute an answer set. The pictures are retrieved by the retrieval system with and without using the taste matrix. The accuracy of the retrieval is evaluated by using precision and recall. Here, precision (recall, respectively) is the ratio of the number of the correct pictures retrieved to the number of the pictures retrieved (the number of all of the correct pictures).

5.2 Evaluation results

(1) Evaluation result of score matrix estimation

The proposed learning method takes 1.5 minutes, whereas the full learning method takes 130 minutes. The learning burden is drastically decreased.

The average scores for *brightness*, *potency*, *sharpness*, *naturalness*, and *activity* of the retrieval are 1.4, 0.4, 0.2, 0.8, and 1.2, respectively. The average scores of all of the factors are more than zero. This means that the retrieval with the proposed personalization is better than that without it.

(2) Evaluation result of factor burden matrix estimation

The evaluation results are shown in Fig. 2 to Fig. 6. These are for the factors: brightness, potency, activity, naturalness, and sharpness, respectively. The number of points in the figures are different from one another. This is because the numbers of the correct pictures are different from one another. The retrieval results with the taste matrix are better than those without it for all of the factors. Especially, this tendency is clear in the factor potency (Fig. 2). Moreover, the retrieval is perfect in the factor sharpness (Fig. 5). That is, no wrong picture is retrieved. These results show that the retrieval could be adapted to the individual by using the taste matrix.

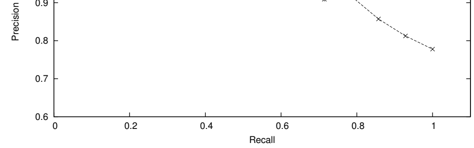

Figure 1. Evaluation result of brightness.

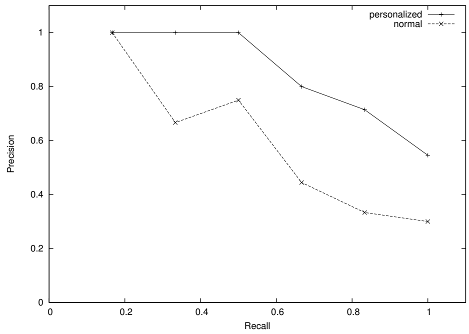

Figure 2. Evaluation result of potency.

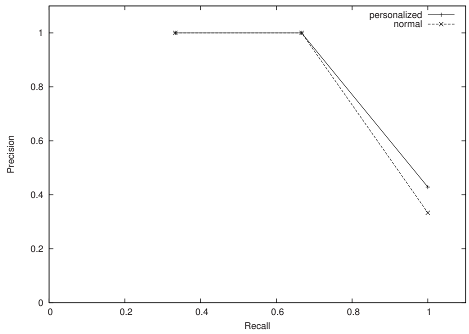

Figure 3. Evaluation result of activity.

5.3 Considerations

(1) Consideration of score matrix estimation

The learning burden can drastically be decreased by using the proposed methods. This will enable users to try the personal adaptation.

The fitness of the retrieval varies according to the factors. This may be caused by the impression word pairs used in the learning. The number of the impression word pairs for brightness is three. That for naturalness is two. Those for potency and for activity are one. There is no impression word pair for sharpness. In the evaluation, the score of the fitness for the factor brightness is best, while that for sharpness is worst. The number of the impression word pairs may relate to the score of the fitness of the retrieval. The consideration of the selection of the impression word pairs is in the future work. As the average score of the fitness of the retrieval is 0.8, personal adaptation can be attained by using the proposed learning method. However, the fitness is evaluated through the range of 3 to -3. The average score 0.8 could not be said to be high. The improvement will be required.

(2) Consideration of factor burden matrix estimation

The evaluation results show that the method using a taste matrix can be used for the personal adaptation. However, the accuracy of the retrieval using the average score matrix is not bad for all of the factors but potency. This means that the retrieval without the personal adaptation can sufficiently be used in usual. When a user strongly requires that the retrieval results follow his/her own taste, the personal adaptation may be in effect.

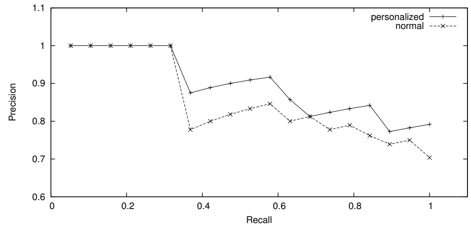

Figure 4. Evaluation result for naturalness.

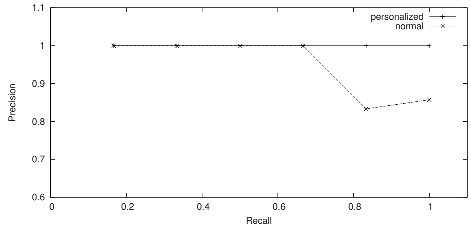

Figure 5. Evaluation result for sharpness.

6. Concluding Remarks

Some researchers may consider that applying the factor analysis for each user is the best way for the personal adaptation. This may be true for the system treating single media data, e.g. picture retrieval system, and music retrieval system. We study on the cross-media retrieval system. In this kind of system, the cooperation of more than one kind of multimedia data is required. The Kansei factors shared with all of the kinds of multimedia data have to be kept. If not, a kind of multimedia data could not correspond to another kind of multimedia data. This results in the stop of the service of the cross-media retrieval. The proposed method using a kind of matrix rotation could keep the Kansei factors. This kind of method is required for the cross-media retrieval system.

Future work includes the improvement of the fitness of the retrieval. Sharing the personality is another issue of the future work.

References

[1] S. Uemura, H. Arisawa, M. Arikawa, and Y. Kiyoki; Digital Media Information Base; IEICE Trans. Inf. & Syst., E82-D(1), 22-33 (1999).

[2] S. Kato and T. Kurita; Visual Interaction with Electronic Art Gallery; Proc. of Database and Expert System Application (DEXA'90), 234-240 (1990).

[3] T. Hochin, and T. Tsuji; Mutual Multimedia Access Using Kansei Factors; Kansei Engineering International, 2(4), 9-18 (2001).

[4] J. G. Snider and C. E. Osgood; Semantic Differential Technique -A Sourcebook; Aldine Pub. Company (1969).