A Method for Constructing a Movie-Selection Support System Based on *Kansei* Engineering

Noriaki Sato, Michiko Anse, and Tsutomu Tabe

Department of Industrial and Systems Engineering, Graduate School of Science and Engineering, Aoyama Gakuin University 5-10-1,Fuchinobe,Sagamihara-shi,Kanagawa-ken, 229-0006, Japan
c5606063@cc.aoyama.ac.jp
{anse, tabe}@ise.aoyama.ac.jp

Abstract. When a person requests, for example, "I want to see a bright and exciting movie," the words "bright" and "exciting" are called *Kansei* keywords. With a retrieval system to retrieve recommended movies using these *Kansei* keywords, a viewer will be able to select movies that fit the *Kansei* without actually having to view samples or previews of the movies. The purpose of this research is to clarify a method to construct a support system capable of selecting movies that fit the viewer's *Kansei,* and to verify the effectiveness of this method based on *Kansei* engineering, for the selection of recommended movies. To accomplish this, we extract the features of a movie using factor-analysis from data from a Semantic Differential Gauge questionnaire, then link the viewer's *Kansei* with the features using multiple linear regression analysis. After constructing a prototype・system to verify the effectiveness, ten examinees viewed a movie selected by the prototype・system. "The selected movie fit the Kansei" at a level of about 70 percent.

Keywords: *Kansei*, retrieval system, factor-analysis, multiple linear regression analysis.

1 Introduction

Ordinarily it is easy to access information through searches based on attributes such as "title," "genre," "performer," etc., using movie-retrieval engines inside conventional rental shops or on websites. Viewers retrieve the movies they wish to see by inputting the keywords via the keyboard. When, on the other hand, a person types in a request such as "I want to see a bright and exciting movie," the words "bright" and "excited" so-called *Kansei* keywords, are difficult to handle. With a retrieval system for recommending movies with registered *Kansei* keywords of this type, the movie viewer will be able to select a movie likely to fit the *Kansei* without actually having to preview the movie. Recent research on opinion extraction has focused on methods to extract opinions from text information such as reviews, book reviews, and text on the Web. Liu, Lieverman, and Selker [2] study ways to assess the feelings of readers who read texts. To understand expressions that evince feelings in a text, methods are used to extract the portions of e-mail text that convey feelings,

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2007, LNCS 4557, pp. 526–534, 2007.

based on categories of feeling defined in advance. Dave, Lawrence, and Pennock [1] classify evaluation expressions from online reviews of products and other online content into the categories of “affirmation” and “denial,” and on that basis extract the opinion. Yu and Hatzivassiloglo [6] use Naïve Bayes classifier to classify newspaper articles into “opinion” and “fact,” then judge whether to classify similar level of sentences and unit of sentence into opinion or not. Moreover, the research focused on text information related to movies deals with subjective expressions extracted from the text. Turney [5] judges whether film reviews are “thumbs up” or “thumbs down.” By keeping track of adjectives from text phrases, they judge whether the phrase is a positive assessment (thumbs up) or negative assessment (thumbs down) based on statistical analysis to quantify commonality. Nakayama and Kando [3] focus on words and phrases from film reviews that indicate feelings, then attempt to understand the characteristic by extracting the “reason” as a related element. The difference of the reason in the same feeling is shown by the writer, and the work of reason is analyzed. Kobayashi [4] focuses on comments from audiences after movie viewings and applies frequency-analysis to connect up with *Kansei* words and phrases categorized as “comments” and “movie genres,” in order to extract *Kansei* words for movies in every genre. It is expected that these-extracted results become effective information for movie selection when shown to users in applications in the field of marketing or work retrieval. However, researchers are further away from accomplishing the same thing in studies on retrieval-method for movies based on subjective expression.

Therefore, we have attempted to clarify a method of movie selection by extracting “human *Kansei*,” that is, subjective expressions, from text information. By relating human *Kansei* with “text,” we construct a retrieval method for movie recommendations that fit two or more *Kansei* requested by viewers. We clarify how to extract human *Kansei* based on the content of a text, and how to statistically analyze commonality in quantitative evaluations of movies and human *Kansei* from an engineering approach. The purposes of this research are therefore to clarify a method for constructing a system for selecting movies that fit a viewer’s *Kansei,* and to verify the effectiveness of our method for constructing a prototype system for selecting recommended movies based on *Kansei* engineering. To accomplish these purposes by three methods: first, we extract the features of a movie using a factor-analysis of questionnaire data with a Semantic Differential Gauge made from *Kansei* keywords; second, we quantify the features of a movement by frequency analysis of keywords associated with the features based on_features present in the movie story which was obtainable easily; third, we analyze how a viewer’s *Kansei* and the features of a movie connect up using multiple linear regression analysis. Through these methods, we construct a prototype support system for a movie-selection engine to retrieve movies that fit a viewer’s *Kansei*, and then test the system by a trial run.

2 Constructing the Support System

2.1 Process Concept

Fig. 1 shows the concept of the support system. Items 1, 2, and 3 in Fig. 1 are the content of the processing.

First, the features of a movie are extracted to clarify the structured factors of the movie.

Second, the features of the movie are quantified by frequency-analysis of the keywords associated with the features present in the movie story, after extracting keywords collected for each feature based on interpretations of factors obtained.

Third, a multiple linear regression formula is drawn up to calculate a user *Kansei*-evaluation value from the questionnaire results.

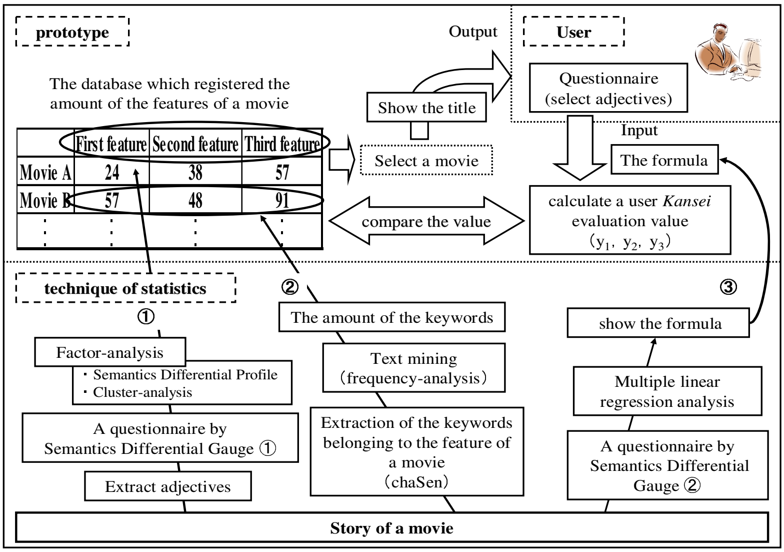

Fig. 1. Concept of the Support System

2.2 Extract the Movie Features

Movie features are collected for the stories of two or more movies, then 20 frequently appearing adjectives (set up uniquely) are extracted from the stories to determine which of the adjectives show the "the *Kansei* feature of the movies." These adjectives are grouped into opposite pairs and are considered in the Semantics Differential Gauge questionnaire in five steps (step 1 to step 5). To investigate the relation between "a movie genre" and "*Kansei* feature" in this case, we have an audience of 20 examinees view a movie preview chosen from among 11 movie genres and then have the examinees answer the above-mentioned questionnaire. Each adjective-pair is estimated by the examinee's *Kansei*. The 11 movie genres chosen are the genres with the highest frequency of appearance in our movie genre investigation.

The factors constituting movies are clarified by a factor analysis of the questionnaire data, then three factors are extracted. Table 1 arranges the adjective pairs based on the factor loads for each in descending order. The feature of the factor is expressed by adjective-pairs over 0.45 points of the absolute value of the amount of

Table 1. Factor Loads

Adjective-pairs		The 1st Factor	The 2nd Factor	The 3rd Factor
very serious	very laughable	0.87	0.14	−0.05
very depressed	very cheerful	0.86	0.27	−0.09
very tragic	very comical	0.85	0.15	−0.03
very dark	very bright	0.84	0.35	−0.15
very sad	very pleasant	0.82	0.18	0.04
very fine	very painful	−0.70	0.15	−0.18
very deep	very light	0.69	−0.23	0.17
very lonely	very busy	0.66	−0.01	0.45
very fearful	very peaceful	0.65	0.41	−0.42
not impressive	very impressive	0.14	0.86	−0.22
not lovely	very lovely	0.20	0.76	−0.21
not heart-warming	very heart-warming	−0.07	0.68	−0.14
very cold	very warm	0.56	0.61	−0.32
very slow	very speedy	0.18	−0.21	0.85
settled down very much	very excited feeling	−0.07	−0.28	0.81
very unreal	very realistic	0.00	0.10	−0.46

each factor-load. The first factor has nine adjective-pairs ("Very laughable," etc.) The second factor has four adjective-pairs ("Very impressive," etc). The third factor has three adjective-pairs ("Very speedy," etc.).

From the above, it turns out that the movie is composed of three obtained factors. Thus, we analyze the meaning of each factor as a key for treating these factors quantitatively. We interpret a factor by two methods: first, we create an SD profile for correlation with the three obtained factors and "movie genre"; second, we reclassify the genres, grouping those that cannot be distinctly divided by the SD profile into "similar impression" categories by cluster analysis (categories that make similar impressions on audiences).

First, we investigate the SD profile by calculating the average value of the questionnaire result for each adjective-pair associated with the three factors for every title. The strength of the factor of a title is taken as the number of adjective-pairs whose average values from five steps are "2 or less and 4 or more." A factor is featured for the highest genre of the rate of the sum total of strength by each of three factors, as each genre has three titles at a time (Refer to Table 2). As a result, the "Comedy" genre belongs to the first factor, the "Romance" genre belongs to the second factor, and the "Horror" genre belongs to the third factor.

Second, we carry out a cluster analysis based on the factor-score obtained by the factor analysis for every genre. As a result, "Horror" from the third factor is positioned in a close relation to "Action" and "Science Fiction/Fantasy." Therefore, it is judged that these movie genres make similar impressions upon viewers. In addition,

Table 2. Factor Strengths

The genre of a movie	The 1st Factor	The 2nd Factor	The 3rd Factor
Animation	18.5	33.3	44.4
Comedy	100.0	33.3	44.4
Action	3.7	0.0	55.6
Youth	3.7	33.3	0.0
Human Drama	3.7	66.7	0.0
Science Fiction/Fintasy	7.4	33.3	55.6
Documentary	0.0	58.3	0.0
Period Drama	0.0	8.3	55.6
Suspense	0.0	0.0	44.4
Romance	29.6	83.3	11.1
Horror	0.0	0.0	66.7

"Youth" is added as a genre of the movie in the second factor. None of the movies in the first factor give impressions of any genre. From these results above, it is determined that "Comedy" belongs to the first factor, "Romance" and "Youth" belong to the second factor, and "Horror," "Action," and "Science Fiction/Fantasy" belonging to the third factor. The interpretation of each factor determines the genre as the first, second, or third feature of the movie.

2.3 Movie Features by Morphological Analysis

Each movie featured is assessed by counting the frequency of keywords belonging to the three abovementioned features in the story of a movie.

2.3.1 Keyword Extraction for Every Feature

The keyword is resolved into morphemes after collecting the stories in six genres belonging to the first, second and third features, in order to extract a keyword for every feature. Next, from the morphemes we extract keywords for three parts of speech which are meaningful as words and phrases, i.e., "Noun," "Noun-General" and "Noun-Adjective verb stem." This leaves words and phrases with high frequency (ten times or more), and those that are common to the six genres are deleted from among them. The extracted words and phrases from the above steps are the keywords for each genre. The first-feature keyword is "Comedy," the second-feature keywords are "Romance" and "Youth," and the third-feature keywords are "Horror," "Action," and "Science Fiction/Fantasy." When we count the number of keywords of each feature, we come up with 120 first-feature keywords, 189 second-feature keywords, and 262 third-feature keywords. We define the "wealth" of the keywords for to each feature keyword category as the number of the keywords extracted in each feature common to all keywords (Refer to Table 3).

Table 3. The wealth of one keyword belonging to each feature

The 1st Feature Comedy	The 2nd Feature Romance/Youth	The 3rd Feature Horror/Action Science Fiction/Fantasy
4.8	2.2	3

2.3.2 Feature of a Movie

The frequency of keywords belonging to the three above-mentioned features in the story of one title is counted using frequency-analysis, one of the text mining techniques. Next, the values for the features for every title are determined by the value obtained by multiplying the frequency of keywords belonging to each feature and the wealth per keyword shown in Table 3. In addition, the length of one story (one title), when printed out, takes up about one sheet of A4 paper.

Table 4. Example movie features

Title name (Genre)	The 1st Feature Comedy	The 2nd Feature Romance Youth	The 3rd Feature Horror/Action Sicence Fiction/Fantasy
A.I. (Sicence Fiction/Fantasy)	86.4	114	154
EYES WIDE SHUT (Human)	72	120	66
ICE AGE (Animation)	144	123	151.8
EXORCIST (Horror)	76.8	45	145.2
SIN・CITY (Action)	14.4	69	110
HITCH	86.4	192	99
MASK2 (Comedy)	268.8	54	140.8

2.4 Linking Human *Kansei* with Movie Features

The same questionnaire described in Section 2.2 is used to link human *Kansei* with movie features. Ten members of an audience (examiners) are asked to fill out a questionnaire made up of nine adjective-pairs with a high factor-load for each factor based on the results of the factor analysis described in Section 2.2. To begin with, by checking strong correlations among the adjective-pairs of the five higher ranks with the high factor loads belonging to the 1st factor, it is judged that there is multicollinearity. For adjustment, we delete four adjective-pairs, e.g., from "very depressed – very cheerful" to "very sad – very pleasant." Moreover, we delete the adjective-pair "not heart-warming – very heart-warming," as an examiner considers these features unconsciously (no person consciously considers whether he or she wants to see a movie which will not remain in his or her heart). Twenty two movies from 11 genres are covered in the questionnaire. Thus, a multiple linear regression analysis is performed by the variable decreasing method using a purpose-variable, the strength of the features for each title, an explanation-variable, and an evaluation of an adjective-pair, in order to show formulas of relations for calculating a

Kansei-evaluation value. If adjustment-R2 becomes 0.5 or more, the accuracy of the regression is good. It is assumed that 0.5 or more base points are fulfilled, and that the model has no variable with multicollinearity. According to the analysis result, all adjustment-R2's for each feature were 0.5 or more. The multiple linear regression formula is shown in (1) ~ (3). In addition, Table 5 shows the adjective-pair of X_1~X_9 used for the explanation-variable

Multiple linear regression formulas obtained with Section 2.4

[Value for the first feature] $y_1=60.1*X_1-42.6*X_2+27.9*X_3+24.8*X_4-45.6*X_5+58.4$ (1)
[Value for the second feature] $y_2=-41.2*X_3+45.9*X_4-34.3*X_6-13.5*X_7+9.2*X_9-22.9$ (2)
[Value for the third feature] $y_3=33.0*X_3+34.5*X_4-37.8*X_6+14.0*X_7-11.0*X_9+26.6$ (3)

Table 5. Adjective-pair used for explanation-variable

	Adjective-pair	
Explanation-value	smallness	largeness
X_1	very serious	very laughable
X_2	very fine	very painful
X_3	very deep	very light
X_4	not impressive	very impressive
X_5	not lovely	very lovely
X_6	very cold	very warm
X_7	very slow	very speedy
X_8	settled down very much	very excited feeling
X_9	very unreal	very realistic

3 Construction of a Prototype and an Experiment

3.1 Construction of a Prototype and Its Concept

The purpose of constructing a prototype is to verify the effectiveness of a method for constructing the above-mentioned support-systems. The concept of a prototype is shown in Fig. 2, and an outline is described below.

"Input *Kansei* information": A user chooses the grade of an adjective-pair that fits the user's *Kansei* according to the questionnaire described in Section 2.4.

"Calculate the *Kansei*-evaluation value": The three obtained strengths of the features determine "a user's *Kansei*-evaluation value," a value which substitutes the evaluation-values of the respective adjectives input for the multiple linear regression formula gained from Section 2.4.

"Database": A database registering the title of a movie and the strengths of the three features is constructed in advance.

"Selection-processing of a movie": The Euclid distance is calculated by the difference of "a user's *Kansei*-evaluation value" and "the strengths of the features of a movie" after comparing a user's *Kansei*-evaluation value with the database. Next, two movie titles are selected from the smallest value.

"Display movie's name": The selected move title is shown to the user.

In addition, the title of a movie in an impartial genre is selected and registered into the database. To verify the effectiveness and to let the user actually see a movie, a movie already burned onto a DVD (DVD-ized) is adopted.

Moreover, the DVD-movie is already adopted because an examiner has actually seen the movie to verify the effectiveness.

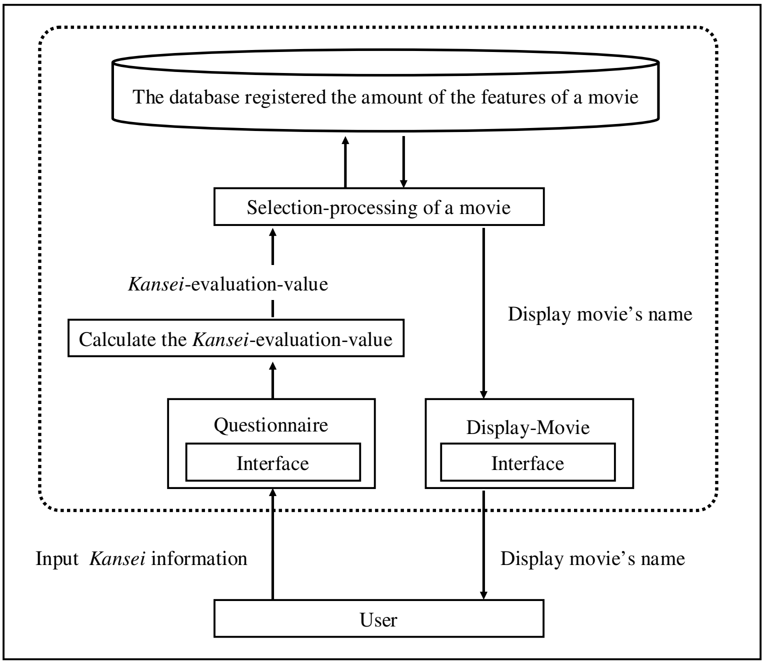

Fig.2. Conceptual Figure of the System

3.2 Verifying the Effectiveness

The experiment is carried out by ten examinees using a prototype. Each examinee chooses an adjective matching the *Kansei* of the movie he or she wants to see. Next, the examinee sees one movie between two other movies selected by the prototype from the results in consideration of the burden of time, that is, the time required to see a movie. Then, when the examinees are asked, in the questionnaire survey, "To what degree (expressed in percent) does the movie you have just seen fit your *Kansei*?," 70% or more of the examinees reply that the selected movie currently fits their *Kansei*.

4 Conclusion

It is clarified that the prototype system built through this research is able to select the movie which most closely fits the *Kansei* a user seeks, among the movies registered in the system database based on verified the effectiveness. Therefore, this research clarifies the method to construct a system capable of selecting movies that fit a viewer's *Kansei*. To raise analytic accuracy in the future, we hope to accomplish the following: first, analyze methods other than correlation with "movie genre" to interpret the factors obtained; second, to examine questionnaire methods that pose less of a burden on examinees, in order to collect more experimental data; third, to automate a series of processes for reading the stories and attributes of movies and registering them into a database.

References

1. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the 12th International World Wide Web Conference (WWW2003), pp. 519–528 (2003)
2. Liu, H., Lieverman, H., Selker, T.: A Model of Textual Affect Sensing using Real-World Knowledge. In: The Proceedings of IUI 2003 (January 12-15, 2003)
3. Norio, N., Noriko, K.: Analysis of Emotion Expression Focusing on Reason. proceedings of the Institute of Electronics, Information and Communication Engineers (IEICE) 105(291), 51–56 (2005)
4. Shinobu, O.: Kansei by Words and Movies Extracting Kansei Using Text Mining. Journal of Japan Society of Kansei Engineering 5(3), 43–47 (2005)
5. Turney, P.D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 417–424 (2002)
6. Yu, H., Hatzivassiloglou, V.: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 129–136 (2003)
7. Toshikatsu, H.: Approach to text mining by EXCEL. Ohmsha, Ltd., pp. 68–69 (2002)