An Automatic Modeling Method of Kansei Evaluation from Product Data Using a CNN Model Expressing the Relationship Between Impressions and Physical Features

Hidemichi Suzuki(✉), Atsuhiro Yamada, Kensuke Tobitani, Sho Hashimoto, and Noriko Nagata

Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo, Japan
{hide4831,nagata}@kwansei.ac.jp

Abstract. In the field of Kansei engineering, the approach is often taken of Kansei evaluation modeling expressing the relationships between physical features and impression of an object. However, in the conventional modeling method, personnel and time costs are very high because multiple experiments and analyses are needed to high precision modeling. In contrast, study using machine learning has been conducted as a method of modeling the relationship between physical features and impressions of products. However, no studies have been reported that considering how the nature of an impression that there is evaluation vary from person to person. In this study, we work on automatically Kansei evaluation modeling using images and review-text data of products existing on the web. A convolutional neural network (CNN) is used for modeling, and variation in the impressions of each product are taken into consideration when learning. In the proposed method, we performed the following: (1) Extraction of the main impressions of target domain and calculation of values that express the strength of each impression from review-text data through text mining based on the previous study [1], (2) creation of a product image data set that uses the distribution of products' impression scores as a training label and (3) construction of the CNN model using the created data set. We applied proposed method to wristwatches as the target domain and verified the estimation accuracy of constructed CNN model. As a result, a high positive correlation was confirmed between estimated impression score and impression scores that were calculated from review-text data. In addition, since present results exceeded the estimation accuracy of CNN model hasn't learned distribution of impression scores, learning variations in the evaluation of peoples' impressions were shown to be effective for improving estimation accuracy.

Keywords: Kansei engineering · Text mining · CNN · Appraisal dictionary

1 Introduction

In the field of product design, it is important to reflect user's needs in products. Particularly in recent years, affective needs such as usability and comfort have attracted attention in addition to conventional manufacturing needs such as function, price and

C. Stephanidis (Ed.): HCII 2019, CCIS 1032, pp. 86–94, 2019.
https://doi.org/10.1007/978-3-030-23522-2_12

reliability [2]. Kansei engineering approach is accepted to be the most reliable and effective method to handle affective needs and is applied to various domains [3]. One specific approach is modeling Kansei evaluation that expresses the relationship between physical features and impressions of a product. With this approach, it is possible to accurately and efficiently reflect user's affective needs on product design.

However, in the conventional method of Kansei evaluation modeling based on a subjective evaluation experiment [4], a subjective evaluation experiment with semantic differential (SD) method and multiple experiments and analyses during the preparation stage are necessary. Therefore, there is a problem in that personnel and time costs are very high.

On the other hand, studies using machine learning have been conducted as a method to model the relationship between physical features and impressions of products has been conducted [5]. However, no studies have been reported that consider how the nature of an impression and its evaluation vary from person to person.

In this study, we work on automatically Kansei evaluation modeling using images and review-text data from many products existing on the Web. In proposed method, based on a previous study [1], the main impression of the target domain (impression topic) is first extracted from whole review-text data via text mining, and then the value that expresses the strength of the impression (impression score) is calculated from some review-text data for each product. Next, product image data set is created with the calculated distribution of the impression scores as training label. Finally, a CNN model that solves estimation of impression scores of product images as classification problem is constructed using created data set. This model solves the estimation of impression scores of product images as classification problem. Figure 1 shows the flow of automatically modeling Kansei evaluation using text mining and a CNN model.

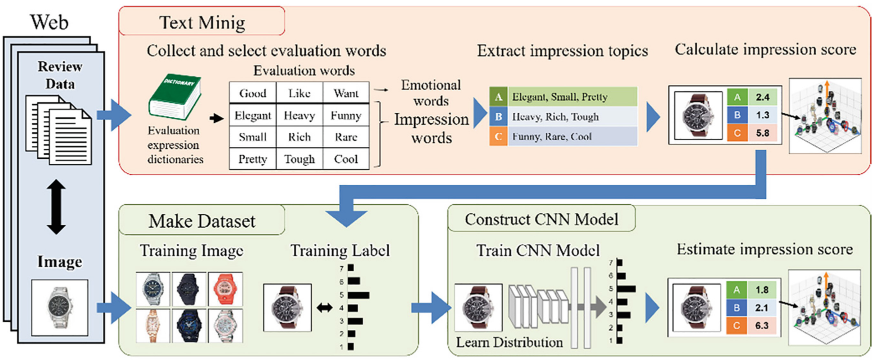

Fig. 1. Flow of the proposed method

2 Method of Extracting Impression Topics and Calculating Impression Scores Through Text Mining

First, evaluation words are collected and classified. At the stage of collecting, affective expressions (e.g. evaluation word candidates such as adjectives) contained in review-text data are collected by consult some evaluation expression dictionaries and part-of-speech information. This candidate evaluation word group compares high-order evaluation words (emotional words) related to a person's emotions and low-order evaluation words (impression words) describing product impressions [6]. However, since the emotional words (such as "want" or "joyful") do not express the features of a product, it is difficult to reflect them directly on the product even if the emotion can be estimated. Therefore, the candidate evaluation word group is classified into emotional words and impression words, and only impression words are used. Classification of evaluation words is performed by consult Japanese evaluation appraisal dictionary [7] in which attributes of internal evaluation and external evaluation are assigned to evaluation expressions. Internal evaluation is expression that indicates evaluator's emotions with respect to evaluation target, or action that represents emotion, an external evaluation is expression that indicates the characteristics of evaluation target. Based on these definition, internal evaluation words are classified as emotional words, and external evaluation words are classified as impression word.

Next, impression topics are extracted. Impression words obtained at previous section are input to HDP-LDA [8], which is a language model that probabilistically finds topics of words in sentences using sentences as input, and extracts impression topics.

Finally, using the frequency of impression words in each review-text data and the importance of each impression word in each impression topic, the impression topic score for the product is calculated. At that time, term-score [9] is used as the importance of each impression word.

In previous study [1], the impression score of multiple review-text data for each product was calculated, and their average value was used as the final impression score of the product.

3 Method for Data Set Creation

3.1 Review-Text Data

First, image and review-text data of products existing on the web are collected. At that time, we set that product with 10 or more review-text data are collected.

3.2 Election of an Impression Topic

Next, impression topics are extracted from review-text data through text mining. At that time, impression topic could include function of product or haptic impression. Such impression topics is considered difficult to estimate from images. Therefore, visual impression topic which is expected to be effective to estimate from images is collected.

3.3 Selection of Product Image

Among the product images collected, there are images of objects and backgrounds other than the product, and those of the target product that cannot be seen. When estimating impression scores from product images, these images are excluded from the data set because they are thought to lead to a reduction in estimation accuracy.

3.4 Creation of Training Label

First, select review-text data. Due to the nature of calculating impression scores method in previous studies [1], the scores calculated from review-text data with fewer impression words tend to be lower than actual scores. Therefore, top 10 review-text data with many impression words are used to calculate impression score of each product.

Next, clustering of impression scores is performed. Since the impression scores calculated from review-text data is continuous value, it is difficult to create a training label of classification problem. Therefore, using k-means method, all impression scores in one impression topic are classified into seven clusters, and class labels from 1 to 7 are assigned in descending order from the impression scores to the product. The reason why clustering number is 7 is to unify with following subjective evaluation experiment.

Finally, probabilities of each classes in which the product will be classified are calculated from 10 class labels of each product and obtained probability distribution is used as training label. This label makes it possible to learn distribution of impression scores and model Kansei evaluation that considers distribution of person's evaluation.

4 Impression Estimation Method of Product Images that Uses the CNN Model

4.1 Architecture of the CNN Model

The network architecture of CNN model used in this method is based on CNN-M [10]. In addition, Table 1 shows parameters that are readjusted from CNN-M for our model.

By learning data set described in this chapter, we constructed CNN model that estimates impression of product image.

Table 1. List of readjusted hyper parameters.

	Number of units in the output layer	Initial value of the learning rate	Normalization method	Initial value of the weight
CNN-M	1,000	10^{-2}	Local response normalization	Gaussian initialization
Proposed method	7	10^{-3}	Batch normalization [11]	He initialization [12]

4.2 Creation of Impression Estimates Using the Constructed CNN Model

The output of the constructed CNN model is a probability distribution of 7 classes. In this method, calculate the correlation coefficient with the impression scores that were calculated from the review-text data through text mining to verify the accuracy of CNN model's impression estimation. We used Eq. 1 to calculate the expected value for each product image from the probability distribution of the output result and used the expected values as the impression scores that the CNN model estimated. In Eq. 1, c_i indicates each evaluation value of 7 classes ($c_1 = 1,\ c_2 = 2, \ldots,\ c_7 = 7$).

$$E|x| = \sum\nolimits_{i=1}^{n} c_i p_i \quad n = 7 \tag{1}$$

5 Creation of Data Set for Wristwatches

5.1 Collected Product Data

The data to be used were collected from Rakuten Market, an online mall that Rakuten, Inc., operate. The total number of wristwatch items was 2,811, and the number of targeted reviews was 252,228. As a result of morphological analysis and classification, 3,880 impression words were collected from the review-text data.

5.2 Selection of Impression Topics and Product Images

The number of extracted impression topics was 9. Table 2 shows each topic. Of these topics, Topic 1 (sophisticated, pretty, small) and Topic 5 (heavy, high grade, nicely textured) were considered related to visual impression.

Then, 1,936 images were selected from 2,811 product images of wristwatch, and training labels were created for data set. Figure 2 shows part of this data set.

Table 2. Extracted impression topics.

Topic number	Impression topics
1	**sophisticated, pretty, precious, tiny**
2	hard to see, dressed, affordable, easy
3	accurate, thin, useful, unnecessary, easy
4	hard, smart, casual, thin
5	**heavy, high grade, durable, Nicely textured**
6	cool, light, hard to see, childish
7	childish, affordable, breakable, enough
8	childish, undisturbing, forgettable, affordable, loose
9	luxurious, fulfilling, overjoyed, unfriendly

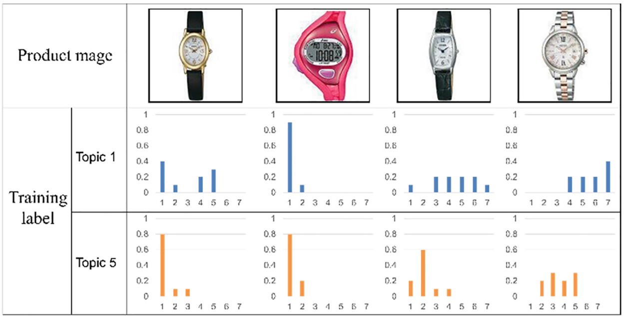

Fig. 2. A part of data set of wristwatches

5.3 Validation of the Impression Scores

To verify the accuracy of impression scores that were calculated based on text mining method [1], a subjective evaluation experiment which answered strength of impression topic feel to product image of data set at 7 stages was conducted. Then, average of 20 people's evaluation value was taken as person's impression score for the product image.

Result

The correlation coefficient between the impression scores calculated from the review-text data and the evaluation scores obtained from the experiment is shown in Table 3. The impression scores calculated for each impression topic is shown on the horizontal axis, and the plots for each impression topic are shown in Fig. 3.

Table 3. Correlation coefficient between average of calculated scores and evaluation scores.

Impression topic	Correlation coefficient
1	0.63
5	0.45

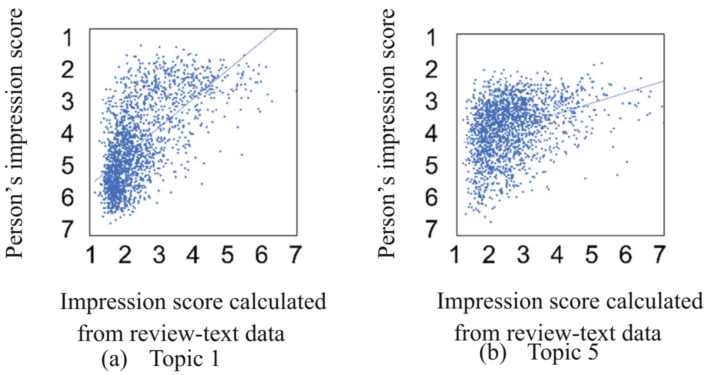

Fig. 3. Plot of the estimation results (text mining)

Discussion

From Table 2, correlation coefficient is 0.63 in topic 1, and strong positive correlation with person's evaluation scores is seen. In topic 5, correlation coefficient is 0.45 and moderate positive correlation with person's evaluation scores is confirmed.

The cause of result that high correlation was not found in topic 5 is considered to be "heavy". It involves not only visual but also tactile impressions, and the reviewer considers both impressions for evaluation. In contrast, it is difficult to estimate tactile impression from image in subjective evaluation experiment. It is thought that such a difference of the situation of evaluation influences correlation with person's evaluation.

6 Impression Estimation of Project Images Using CNN Model

In this study, we construct CNN model that estimates the scores of impression topic 1 (sophisticated, pretty, small), which has higher correlation with the person's evaluate scores. For accuracy verification, K-fold cross validation with K = 11 was performed.

In addition, to investigate the effect of using distribution of impression scores for training label on estimation accuracy, we created training label which does not consider distribution of impression scores. In procedure, first, an average value of impression scores calculated from review-text data is obtained for each product. Next, with k-means method, average scores are classified into 7 classes, and class labels from 1 to 7 are assigned as in Sect. 3. Finally, create an array whose element of class label number is 1 and the other is 0 (if the class label is 3, the array is [0, 0, 1, 0, 0, 0, 0]). Using this array as training label, we constructed CNN model with procedure described in Sect. 4 and compared with proposed method.

6.1 Result

The correlation coefficient between the estimation results obtained by the proposed method and the impression score calculated from the results obtained by the comparative method is shown in Table 4. In the proposed method and comparison method, the expected value calculated using Eq. 1 was used as the estimation result Fig. 4 shows a plot of the estimation results of each method.

Table 4. Difference of the training labels and the correlation coefficient

	Training label	Correlation coefficient
Proposed method	Impression distribution	0.67
Comparison method	Average of impression score	0.51

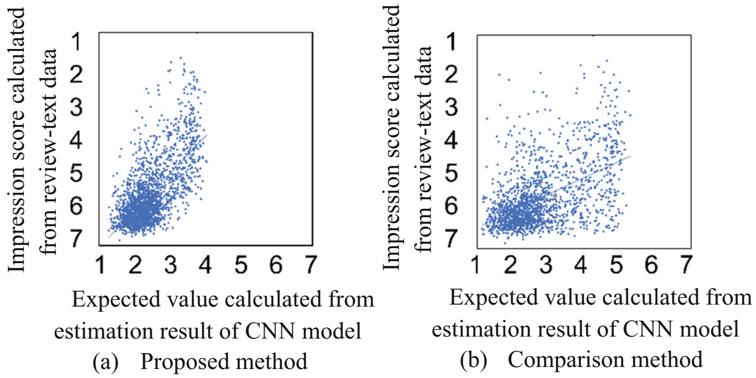

Fig. 4. Plot of estimated result (CNN)

6.2 Discussion

From Table 3, we confirmed positive correlation with correlation coefficient of 0.67 between expected value with proposed method and impression score calculated from review-text data, and effectiveness of proposed method was confirmed. In addition, since this result exceeds correlation coefficient 0.54 calculated with CNN model which hasn't learned distribution of impression scores, it was shown that learning distribution of impression evaluation is effective for improving estimation accuracy.

7 Conclusion

We created image data set considering impression distribution of images by collecting product images and review-text data on the web and using text mining. In addition, we proposed a method to automatically modeling Kansei evaluation that estimates impression from image by constructing CNN model using created data set.

In addition, effectiveness of proposed method was verified with wristwatch as target product. As a result, we confirmed high positive correlation between impression scores of training data and estimated impression scores, and confirmed effectiveness of method. Furthermore, this result exceeds estimation result of CNN model hasn't learned distribution of impression, and it was shown that learning distribution of person's impression evaluation is effective for improving estimation accuracy.

References

1. Hashimoto, S., Yamada, A., Nagata, N.: A quantification method of composite impression of products by externalized evaluation words of the appraisal dictionary with review text data. Int. J. Affect. Eng. **18**(2), 59–65 (2019)
2. Toyoda, N., et al.: Objective evaluation about texture for cosmetic ingredients by direct shear testing of powder bed. J. Soc. Pow. Technol. **52**, 694–700 (2015). https://doi.org/10.4164/sptj.52.694

3. Chen, C.H., Khoo, L.P., Chen, K., Pang, J.H., Huang, Y.: Consumer-oriented product form creation via Kansei engineering. In: Proceedings of the International Symposium for Emotion and Sensibility e Emotion Research in Practice, pp. 184–191 (2008)
4. Tobitani, K., Matsumoto, T., Tani, Y., Fujii, H., Nagata, N.: Modeling of the relation between impression and physical characteristics on representation of skin surface quality. J. Inst. Image Inf. Telev. Eng. **71**, 259–268 (2017)
5. Lu, X., Lin, Z., Jin, H., Yang, J., Wand, J.Z.: Rapid: rating pictorial aesthetics using deep learning. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 457–466. ACM (2014)
6. Sano, M.: "Nihongo apureizaru hyooka hyoogen jisho -taido hyooka hen-" no koochiku ∼ hyooka no tayousei wo toraeru tameno gengoshigen no kaihatsu ∼. In: Proceedings of the Annual Meeting of the Association for Natural Language Processing, vol. 17, p. ROMBUN NO.E1-2 (2011)
7. Sano, M.: Japanese Dictionary of Appraisal -attitude-. Gengo Shigen Kyokai (2010)
8. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Sharing clusters among related groups: hierarchical Dirichlet processes. In: Advances in Neural Information Processing Systems, pp. 1385–1392 (2005)
9. Blei, D.M., Laffety, J.D.: Topic models. In: Text Mining: Classification, Clustering, and Applications, vol. 10, p. 34 (2009)
10. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. CoRR, abs/1405.3531 (2014)
11. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. The IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)