

A Method of Music Retrieval Using Impression Keywords

Kazuhiro Morita*, Masao Fuketa, Jun-Ichi Aoe

Institute of Technology and Science, The University of Tokushima, Japan

* Corresponding author. Email: kam@is.tokushima-u.ac.jp.

Manuscript submitted August 17, 2014; accepted June 8, 2015.

doi: 10.17706/jsw.10.6.775-783

Abstract: Recently, making electronic data of the multimedia such as music, images, books, and so on has been spread. However, it is necessary to look for specific contents in the database with a large amount of music data because electronic data increases rapidly. Therefore, an efficient search engine is requested. This paper presents a music retrieval method focusing on the impression keywords. In the presented method, impression keywords as queries is constructed by using Twitter beforehand. The impression vector to music is generated by using this impression keyword set. And, the retrieval is processed by measuring the similarity to impression vectors with a query. From experimental results, it turns out that the recommended music displayed as a retrieval result appropriately expresses the impression of the user requests.

Key words: Impression keyword, music retrieval, similarity measurement.

1. Introduction

Recently, making electronic data of the multimedia such as music, images, books, and so on is advanced as the telecommunication technology develops. As a result, anyone is able to acquire various information easily by way of the Internet. Especially, the music data can be used by downloading directly from an online shop for electronic equipment such as a personal computer, smart phones, and portable music players. However, it is necessary to look for specific contents in the database with a large amount of music data because information accumulated as electronic data increases rapidly. Therefore, an efficient search engine is requested.

In general, the metadata of music such as the title, songwriter, composer, lyrics and so on is used for the retrieval of the music data. However, when the metadata cannot be used if the user doesn't know the title, this retrieval method cannot be applied. Therefore, there are studies related to the annotation[1][2] and the tagging [3], [4] in order to retrieve easily by enhancing the metadata. Moreover, there are researches which target music is retrieved by humming a part of the tune without using the metadata[5].

However, only using the metadata is difficult to consider the impression to the music without listening. Therefore, there is a possibility that the audition and the review confirmation are needed whether music is suitable for the image that the user requests or not, and the retrieval efficiency gets worse consequentially.

The technique for the sensibility retrieval based on the impression of the music by the user requests is enumerated as one of the methods in order to solve such a problem. The sensibility retrieval is a method to retrieve multimedia data by using the word that shows an arbitrary impression based on user's some criteria such as a preference and a concern.

In the related researches on the music retrieval using the sensibility [6], [7], the audition experiment of

the music data using the semantic differential scale method is conducted at first. The feature of music is generated from the result of the audition experiment with the technique of multivariate statistics. In the retrieval, the music suitable for the feature is retrieved by giving each index of the semantic differential scale. Therefore, the index input by the retrieval complexly becomes seven levels and 8-10 kinds.

To solve this problem, the music retrieval system by the natural language dialogue is proposed [8]. This system prepares the impression word often used by the music retrieval, and associates the impression word with the index of the semantic differential scale by the experiment. Also, a natural retrieval is conducted by extracting the impression words from the input sentence, and generating the input of the music retrieval from the index corresponding to the impression words.

However, it is difficult for these techniques to expand music newly, because the audition experiment is necessary for making the feature of music. Moreover, the problem is how to expand the number of words and how to pick up the impression words in the retrieval by the natural language. Therefore, it is thought that there are a lot of insufficient points for the contemporary music which diversity is increased.

This paper presents the music retrieval method focusing on the impression keywords. In the presented method, impression keywords as queries is constructed beforehand. The impression vector to music is generated by using this impression keyword set. Also, the retrieval is processed by measuring the similarity to impression vectors with a query. Moreover, Twitter that can acquire latest information at any time is used in order to construct the impression keywords and to generate the impression vectors. From experimental results, it turns out that the recommended music displayed as a retrieval result appropriately expresses the impression that the user requests.

2. Construction of Impression Keyword Database

2.1. Review Stage

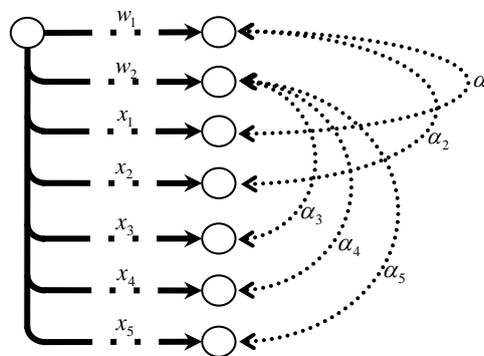


Fig. 1. Outline of the link trie structure.

In this paper, the keyword that shows an arbitrary impression is defined as an impression keyword. Impression keywords are various and fluidly because there are expressions which are felt from objects by users. Then, the impression word database needed in the music retrieval method is constructed at first.

The construction process is as follows;

Step 1: Tweets including keyword $w(\in W)$ is retrieved from Twitter, where W is a keyword set registered in the impression keyword database.

Step 2: From retrieved tweets, related words $x_k(1 \leq k \leq u)$ for w are acquired by using the morphological analysis and the pattern matching in each sentence. The acquired related words x_k are registered in the database as new keywords (i.e. $x_k \in W$). At this time, related information α_k between w and x_k is registered.

As for the acquisition of x_k , the study related to acquisition for the sensibility knowledge[9] is applied. This construction process is able to collect a lot of impression keywords by repeating newly registered keywords. Moreover, it is also possible to correspond to the change of the impression which users feel by using Twitter.

When let (w, x_k, α_k) be the triplet for related information α_k between w and x_k , (w, x_k, α_k) is registered in the impression keyword database. Furthermore, w and x_k should be able to be retrieved individually. Then, the Link Trie Structure that Morita et al.[10] proposed is used for implementing the impression keyword database. Fig. 1 shows the outline of the Link Trie Structure. Because the Link Trie Structure considers w and x_k to be an independent key, x_k can be registered as a new word and the relation of the triplet (w, x_k, α_k) is maintained

Table 1. Example of Registered Words

Adjective	1st acquisition	2nd acquisition
pretty	cute	beauty
		fancy
		Pop
	Kawaii	favorite
		healing
	warm fuzzy	leisurely
		relaxed
		heartwarming
	little	mini
		micro
		S

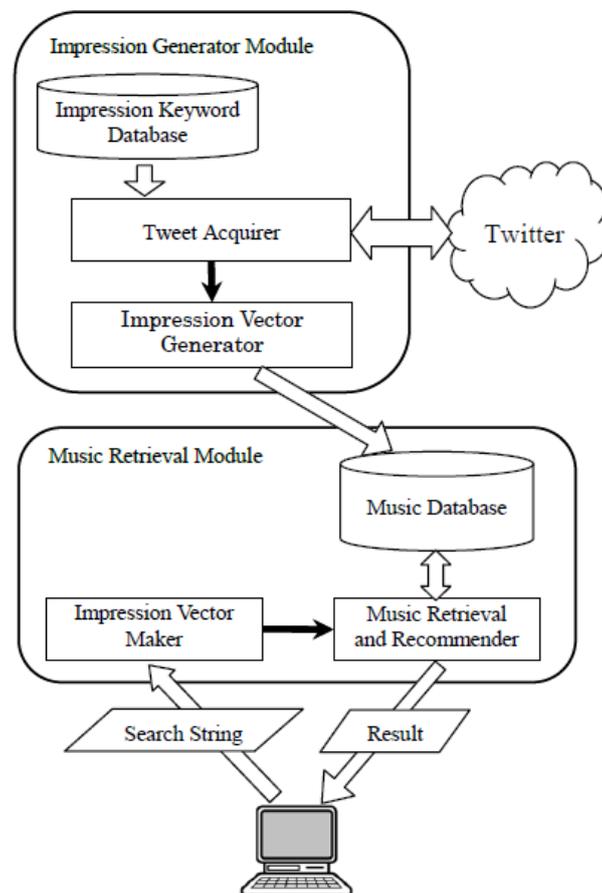


Fig. 2. Flow of the presented music retrieval method.

In this paper, ten adjectives (funny, pretty, happy, sad, beautiful, fearful, wise, nice, light and heavy) are manually registered in the impression database as first keywords, and construction begins. Also, 165 keywords are finally registered by repeating twice, and this database is used for the presented music retrieval method. Table 1 shows the example of the registered keywords. In the impression keyword, abbreviations and variants are allowed. Even if one word has one meaning, each user might have different impression. In addition, onomatopoeia is registered for similar reasons.

3. Music Retrieval Method

Fig. 2 illustrates the flow of the presented music retrieval method. The proposal method is composed of the impression generator module where the impressions for music are given and the music retrieval module where the constructed database is retrieved by inputting a natural language sentence.

3.1. Impression Generator Module

The presented method gives the impression vector to each music by using the impression keywords collected beforehand, and uses the impression vector for the retrieval. In the impression generator module, the impression vectors are generated and registered into the music database.

3.1.1. Tweet acquisition by impression keywords and the music title

Music title $m(\in M)$ and impression keyword $w_i \in W(1 \leq i \leq n)$ are retrieved from Twitter, and number C_i of tweets is acquired, where M is a set of music titles. The maximum value of C_i becomes 100 from upper bounds of the retrieval number. In the retrieval, bots and periodic tweets are excluded by using the request parameter of Twitter API. The reason is tweets which are not the remarks of contributors are unnecessary.

The reason to use Twitter is to receive in real time latest information from a lot of users. Moreover, it is necessary to bring sentences together concisely to pass on the opinion and the preference because there is a limitation of 140 characters or less. Therefore, it is one of the reasons that frank opinions are often written.

3.1.2. Generation of the impression vectors

The impression vector is defined as a vector quantified the degree of each impression keyword of the impression evoked from an object. The number of tweets is used for the degree of the impression in this paper.

Impression vector V_m in which all C_i acquired by retrieving all w_i and m are as components is generated.

When let R^n be n -dimensional vector space, V_m is defined as follows;

$$V_m = \sum_{i=1}^n c_i e_{w_i} \quad (1)$$

where e_{w_i} is a standard basis corresponding to w_i .

Impression vectors are generated to all music titles and stored into the music database.

The concept of the impression vector applied the contrivance of the word vector in the word clustering technique by using the co-occurrence information [11], [12].

3.2. Music Retrieval Module

This module retrieves suitable music for the input impression by using the impression vector.

3.2.1. Making the impression vector from the search string

First of all, the expression that shows the impression is input as a search string. The nouns, the adjectives, the verbs, and the adverbs are detected by the morphological analysis for this search string. Let S be a set of detected words. Impression vector V_S is made by using words included in the impression keyword database among detected words.

V_S is defined as follows;

$$V_S = \sum_{i=1}^n s_i e_{w_i} \quad (2)$$

$$\text{where } s_i = \begin{cases} 100 & \text{if } w_i \in S \\ 0 & \text{if } w_i \notin S \end{cases}$$

The value of the dimension is 100 according to the maximum value of each dimension of impression vector V_m .

3.2.2. Retrieval and recommendation in the database of music

The similarity of impression vector V_{m_j} and V_S to music $m_j \in M(1 \leq j \leq t)$ in the music database is measured. The cosine distance is used for the calculation of the similarity.

The cosine distance $Dc(V_{m_j}, V_S)$ can be defined as follows;

$$Dc(V_{m_j}, V_S) = \frac{V_{m_j} \cdot V_S}{\|V_{m_j}\| \|V_S\|} \quad (3)$$

The similarity is measured for all music in the database, and three pieces of high ranks of the similarity are displayed as recommended music. Moreover, the following five pieces are displayed as semi-recommended music when there is no music that the user requests from the recommended music.

4. Experimental Results

To confirm the effectiveness of the presented method, the evaluation experiments were performed for the impression generator module and music retrieval module.

4.1. Experiment for Impression Generator Module

The impression vector in this paper is assumed to have the relationships between music and the impression keyword in the tweet. Then, the experiment that evaluates the relation in the tweet is performed.

Ten pieces of pop music are chosen from the ranking of the online music distribution site at random, and the contents of the tweets acquired when the impression vector is generated to those ten pieces is confirmed by hand.

When the impression keyword included in the tweet co-occurs with the music title, it is judged to have the relationships between music and the impression keyword. This relationship is assumed to be a correct answer and accuracy is measured.

The accuracy uses the following expression;

$$\text{accuracy}(\%) = \frac{\text{number of correct answers}}{\text{number of acquired tweets}} \times 100$$

Table 2 shows the result of the experiment. The number of total correct answers was 1453 in 1745, and the accuracy became 83.3%. Accuracies of eight out of ten pieces exceeded 80% though the difference of accuracy of each music was seen. Moreover, 96.4% accuracy which is considerably good result was obtained

in the music with the best accuracy. Therefore, it can be said to have the relationship between music and the impression keyword in the tweet, and is thought that the processing of the impression generation module is appropriate.

For consideration of the result, the music title was used as a brand name, and it was used in the metaphor for the tweet classified into the failure occasionally. It is thought that this problem is improvable to introduce the dependency analysis and to exclude an unnecessary tweet.

Table 2. Accuracy for the Impression Generating

Music no.	Number of acquired tweets	Number of correct answers	Accuracy (%)
1	189	163	86.2
2	245	180	73.5
3	133	123	92.5
4	419	354	84.5
5	114	102	89.5
6	89	81	91.0
7	249	169	67.9
8	107	94	87.9
9	139	134	96.4
10	61	53	86.9
Total	1745	1435	83.3

Table 3. Rate of the Target Music found at the 1st Retrieval

Number of retrieval	Number of target music found at the 1st retrieval	Rate (%)
51	33	64.7

4.2. Experiment for Music Retrieval Module

The music database storing 50 pieces from which the impression vector is given by the impression generation module is used for the experiment. 11 participants in the experiment used the music retrieval module, and the music of the retrieval result was previewed actually. Afterwards, the questionnaire concerning the evaluation of the presented method was taken. The following three items were measured by this experiment, and the accuracy of the music retrieval module was investigated;

- Accuracy of retrieval
- Evaluation of retrieval result
- Effectiveness of music retrieval module

The accuracy of the retrieval shows the ratio which the participant found target music at the first retrieval. Table 3 shows the result.

The accuracy of the retrieval was 64.7%. The reason is that the unregistered keywords and music in the databases were retrieved. The amount of registration of the databases was small-scale in this experiment. Therefore, if the scale is increased, this problem can be solved.

Next, the evaluation of the retrieval result and the effectiveness of the music retrieval module are evaluated by the questionnaire. The evaluation of the retrieval result was judged by using 5-point scale whether the music actually auditioned was suitable for the image of the retrieval keyword or not. For the effectiveness of the music retrieval module, the participant judged the effectiveness and the satisfaction rating the module by a scale of one to five. The evaluation results are shown in Table 4 and Table 5.

Table 4. Degree of Satisfaction for the Recommended Music

Scale of evaluation	Number of music	Rate (%)
5	14	42.4
4	13	39.4
3	3	9.1
2	2	6.1
1	1	3.0

Score	4.12
-------	------

Table 5. Degree of Satisfaction for the Music Retrieval Module

Scale of evaluation	Number of participants	Rate (%)
5	3	27.3
4	7	63.6
3	1	9.1
2	0	0.0
1	0	0.0

Score	4.18
-------	------

It is thought that either result can exceed 4 points, and the great result was able to be obtained. For the evaluation of the retrieval result, it is judged that most of participants were able to find target music from the recommended music that is the first result. However, there was an opinion that suitable music in the semi-recommended music was able to be found. It is thought that the reason is that there is no difference in the similarity of each music when the similarity was measured. The solution of this problem is to enhance the impression keyword when impression vector V_s is made for the search string.

For the effectiveness of music retrieval module, the evaluation of four points or more from 10 out of 11 is obtained, and it can be said that most of participants are satisfied with the module. However, there was an opinion that it needs more retrieval results to be displayed. Although the recommended music is displayed at two stages in the presented technique, it is necessary to examine another method of the display.

5. Conclusion

This paper presented the music retrieval method focusing on the impression keywords. The presented method was composed of the impression generator module where the impressions for music were given from Twitter and the music retrieval module where the constructed music database was retrieved by inputting a natural language sentence. The complexity in the conventional method when the music feature was generated was reduced by using Twitter in order to generate the impression, and it was able to correspond to the latest music easily. From the experiments, the satisfaction rating four or more to the recommended music was 81.8% and excellent results were obtained.

The future work is an improvement of the retrieval efficiency by the expansion of the impression keyword database and the music database, and is an examination for the method of the display.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Number 24700209.

References

- [1] Knees, P., & Schedl, M. (2011). Towards semantic music information extraction from the web using rule patterns and supervised learning. *Music Recommendation and Discovery*.
- [2] Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. R. G. (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16(2), 467-476.
- [3] Symeonidis, P., Ruxanda, M., Nanopoulos, A., & Manolopoulos, Y. (2008). Ternary semantic analysis of social tags for personalized music recommendation. *Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 219-224).
- [4] Tingle, D., Kim, Y., & Turnbull, D. (2010). Exploring automatic music annotation with acoustically-objective tags. *Proceedings of the International Conference on Multimedia Information Retrieval* (pp. 55-62)
- [5] Dannenberg, R., Birmingham, W., Pardo, B., Hu, N., Meek, C., & Tzanetakis, G. (2007). A comparative evaluation of search techniques for query-by-humming using the MUSART testbed. *Journal the American Society for Information Science and Technology*, 58(5), 687-701.
- [6] Ikezoe, T., Kajikawa, Y., & Nomura, Y. (2001). Music database retrieval system with sensitivity words using music sensitivity space. *Trans. IPS of Japan*, 42(12), 3201-3212.
- [7] Kumamoto, T. (2005). Design and evaluation of a music retrieval scheme that adapts to the user's impressions. *Proceedings of the 10th international conference on User Modeling* (pp. 286-296)
- [8] Kumamoto, T. (2007). A natural language dialogue system for impression-based music-retrieval. *Proceedings of the CICLING-07 Computational Linguistics and Intelligent Text Processing* (pp. 19-24) .
- [9] Yoshinari, T., Atlam, E. S., Morita, K., Kiyoi, K., & Aoe, J. (2008). Automatic acquisition for sensibility knowledge using co-occurrence relation. *International Journal Computer Applications in Technology*, 218-225.
- [10] Morita, K., Koyama, M., Fuketa, M., & Aoe, J. (1999). A link trie structure of storing multiple attribute relationships for natural language dictionaries. *International Journal Computer Mathematics*, 72(4), 463-476.
- [11] Morita, K., Kitagawa, H., Fuketa, M., & Aoe, J. (2013). An incremental construction method of a large-scale thesaurus using co-occurrence information. *International Journal Computer Applications in Technology*, 48(2), 120-129.
- [12] Morita, K., Atlam, E. S., Fuketa, M., Tsuda, K., Oono, M., & Aoe, J. (2004). Word classification and hierarchy using co-occurrence word information. *International Journal Information Processing and Management*, 40(6), 957-972.



Kazuhiro Morita is an associate professor at the University of Tokushima, where he was a research assistant from 2000 to 2006 in information science and intelligent systems. He is a member of the Information Processing Society of Japan and IEEE Computer Society. He received his BSc, MSc and PhD in Information Science and Intelligent Systems from the University of Tokushima, Japan in 1995, 1997 and 2000, respectively. His research interests are sentence retrieval from huge text databases, double-array structure and binary search trees.



Masao Fuketa received his BSc, MSc and PhD in information science and intelligent systems from University of Tokushima, Japan, in 1993, 1995 and 1998. He has been a research assistant from 1998 to 2000 in information science and intelligent systems from University of Tokushima, Japan. He is currently a research associate in the Department of Information Science and Intelligent Systems, Tokushima University, Japan. He is a member

in the Information Processing Society in Japan and the association for Natural Language Processing of Japan. His research interests are in sentence retrieval from huge text databases and morphological analysis.



Jun-Ichi Aoe received his BSc and MS in electronic engineering from the University of Tokushima, Japan, in 1974 and 1976, respectively, and PhD in communication engineering from the University of Osaka, Japan in 1980. Since 1976, he has been with the University of Tokushima. He is currently a professor in the Department of Information Science and Intelligent Systems, Tokushima University, Japan. His research interest is in natural language processing, a shift-search strategy for interleaved LR parsing, robust method for understanding NL interface commands in an intelligent command interpreter, and trie

compaction algorithms for large key sets. He was the editor of the Computer Algorithm Series of the IEEE Computer Society Press. He is a member in the Association for Computing Machinery, the Association for the Natural Language Processing of Japan.