Emotion-Based Music Information Retrieval Using Lyrics

Akihiro Ogino, Yuko Yamashita

Emotion-Based Music Information Retrieval using Lyrics

Akihiro Ogino and Yuko Yamashita

Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-City 603-8555, Japan
ogino@cse.kyoto-su.ac.jp

Abstract. In this paper, we present a study on emotion-based music information retrieval using lyrics information. Listeners want to search the lyrics of music suitable for his/her emotion (impression of music), by using an information system from music libraries. As a solution of listeners' needs, we have designed a system that retrieve the lyrics of music based on the emotion (or the impression) suitable for a listener's feelings that the listener has selected, from 9 emotions and 9 impressions. We select the words, i.e. verb and adjective, from the bridge part of the lyrics of music that express emotion in lyrics by using natural language processing. We summarize the words into the representative words by using a dictionary of synonyms. We make a model that estimates a listener's 9 emotion/impression of the representative words by using a machine learning method. And listeners want to understand why the recommended music by a system is suitable for his/her emotion/impression. Therefore, we select the representative words most related to a listener's emotion/impression and we use the selected words as the explanation of reason to a listener. We have made each model of emotion and impression for 9 subjects and have evaluated the accuracy of the model. We also have investigated the selected representative words related to emotion/impression.

Keywords. Affective computing, Music information retrieval, Text Mining

1 Introduction

In order to retrieve the music suitable for listener's emotion/impression from the digital music libraries that store tremendous music, we propose a method of emotion/impression-based music retrieval using lyrics for Japanese music.

As listeners' behaviors for searching music, Lee [1] shows that listeners search the music by the bibliographic information of music, such as singer, title and genre. The query by bibliographic information is useful in when he/she has the accurate information of the targeted music. However, when he/she searches new music, it is difficult to find the music that he/she wants to listen because he/she does not have the information of music. As listeners' important factor for choosing music, the survey also shows that he/she pays attention to the contents of music, such as the lyrics of music and the relationship with emotional state of them, when he/she searches music. A listener understands the meaning of the lyrics of music and selects the music suitable for his/her current feelings or moods.

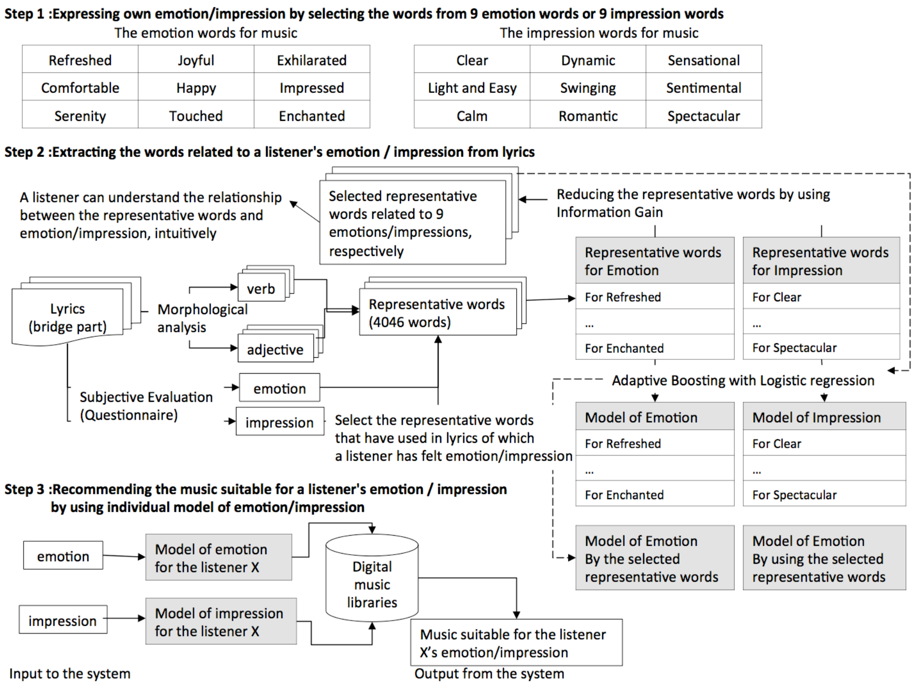

Fig. 1. The framework of our approach for recommending the music suitable for a listener's emotion/impression

Therefore, detecting the listener's emotion/impression of music through analyzing lyrics is useful for a listener to search music that touches his/her minds. There are a number of researches on music retrieval based on emotion/impression by using lyrics [2, 3, 4, 5, 6, 7, 8].

On the other hand, a listener checks the news, reviews and etc. in Internet as a reference because they want to get reasons to select the music. Therefore, for a listener, we think that a music information retrieval system has to have a mechanism that shows the candidate music with the reason.

We propose an approach that estimates a listener's emotion/impression of the music by analyzing the relationship between the lyrics of it and the listener's emotion/impression. We offer the music suitable for his/her emotion/impression with the selection reasons, which are the words of the lyrics more related to emotion/impression, of the music. The framework of our approach is shown Figure 1. It consists of three sections. The first section is the definition of the emotional words that show a listener's emotion/impression. We have selected 9 words for expressing emotion and 9 words for expressing impression in order to support queries by emotion/impression that he/she feels from music. The second section is to extract the words of lyrics related to a listener's emotion/impression.

Table 1. Words for expressing emotion and impression of lyrics of music

Emotion words for music		
Refreshed	Joyful	Exhilarated
Comfortable	Happy	Impressed
Serenity	Touched	Enchanted

Impression words for music		
Clear	Dynamic	Sensational
Light and Easy	Swinging	Sentimental
Calm	Romantic	Spectacular

In order to estimate a listener's emotion of lyrics, we have used the lyrics of the bridge part of music because the bridge has a high likelihood that a listener feels emotion/impression from lyrics in Japanese pop songs. We have inquired a listener's emotion/impression of the lyrics in the bridge part through subjective evaluation test. We have made a model of the relation between the listener's emotion/impression words and the lyrics using a machine learning method.

The third section is to retrieve the music suitable for a listener's emotion/impression from the digital music libraries by using the model automatically. Our system searches the music suitable for the listener's emotion/impression and offers the music to a listener with the words related to the listener's emotion/impression.

The rest of this paper is organized as follows. Section 2 describes a method of emotion-based music retrieval using lyrics. Section 3 presents the evaluation of the proposed method. Section 4 is conclusion of our work.

2 Estimating listener's emotion/impression of lyrics

2.1 Definition of emotional words

We assume that there are the needs of two types in emotional music search: one is a search by emotion and another is a search by impression. For example, one case is that a listener wants to search the music that changes his/her emotion to "Happy", directly. In this case, a listener has clearly the needs, such as "I want to change my sad emotion to happy". Another case is that a listener wants to search the music that offers "Uplifting impression" to his/her mind and to change his/her mind to positive. In this case, a listener's needs are ambiguous and his/her emotion may change to happy or exhilarated. Therefor, we have selected 9 words for expressing emotion and 9 words for expressing impression in order to support queries by emotion/impression that he/she feels from music as shown Table 1.

We select the emotion words from positive words of Russell's model and add words for expressing music, such as "Impressed". We select the impression words that are likely to cause emotion of the emotion words by authors.

2.2 Detecting the words related to emotion/impression from lyrics

We pay attention to adjective and intransitive verb to extract a listener's emotion/impression of lyrics. For example, a human laughs when he/she feels happy. The "laugh" which is intransitive verb expresses happy feelings. We use an adjective word, such as cheerful, when we want to express a happy feeling in the writing. The word of "cheerful" which is adjective expresses happy feelings of a human in the writing. We do not use noun because noun is used to express the surroundings where we are, such as summer, park and sea.

The detecting of the words related to a listener's emotion/impression is composed of four steps:

1. Extracting adjective and intransitive verb from lyrics;
2. Summarizing the extracted words to the representative word based on synonym dictionary of Japanese.
3. Vectorizing of an appearance of representative words.
4. Selecting the representative words related to emotion / impression by using the criteria of information gain

First, we carry out word segmentation of lyrics with the help of natural language processing (NLP) tool for Japanese, i.e. Mecab[11]. We select the adjective and the intransitive verb from the segmented words. Second, we summarize the extracted words to the representative word based on the synonym dictionary. The representative words are 4626 words that are composed by noun, adjective and verb. For example, a representative word includes 67 words. Third, we vectorize the representative words based on an appearance of the word. We set a true flag when a representative word includes an adjective and intransitive verb. We also set a false flag when a representative word does not include an adjective and intransitive verb. Four, we select the representative words, which are more related to emotion/impression than other words, by using the criteria of information gain. We have selected the representative words of which the information gain is over zero.

2.3 Making a model for estimating emotion/impression

We make a model that estimates a listener's emotion/impression of lyrics of music by using a machine learning method. As a machine learning method, we have selected the adaptive boosting with the decision stump as weak classifier. We make one model for one emotion / impression and we make 18 models (9 models of emotion and 9 models of impression) for a listener. For example, when a listener wants to listen to the music of lyrics that affords happy feelings to the listener, we select the model of "happy" from 18 models and retrieve the music of the lyrics from a music library by using the model of "happy".

Table 2. The accuracy rates of models and the number of which subjects have felt emotions, as for Refreshed, Exhilarated and Comfortable (All sample music are 79).

	Refreshed		Exhilarated		Comfortable	
Subjects	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate
A	1	98.7	4	86.1	3	92.4
B	5	84.8	2	97.5	2	94.9
C	12	74.7	16	55.7	18	62.0
D	11	75.9	3	92.4	9	86.1
E	9	78.5	0	---	6	83.5
F	9	83.5	8	78.5	14	70.9
G	13	77.2	14	65.8	11	67.1
H	5	83.5	1	98.7	2	89.9
I	0	---	3	94.9	7	73.4
Baseline	16	48.1	11	86.1	17	58.2

Table 3. The accuracy rates of models and the number of which subjects have felt emotions, as for Joyful, Impressed, Touched and Enchanted (All sample music are 79).

	Joyful		Impressed		Touched		Enchanted	
Subjects	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate
A	6	82.3	35	51.9	14	55.7	14	60.8
B	28	50.6	29	58.2	5	87.3	7	79.7
C	18	62.0	32	46.8	28	63.3	28	43.0
D	11	84.8	37	45.6	16	65.8	15	55.7
E	20	50.6	31	50.6	17	64.6	26	54.4
F	19	58.2	9	83.5	15	77.2	30	57.0
G	17	68.4	16	68.4	14	65.8	5	68.4
H	12	81.0	18	55.7	14	70.9	23	65.8
I	15	68.4	30	59.5	13	73.4	14	77.2
Baseline	36	48.1	52	43.0	32	58.2	49	50.6

3 Evaluation

In order to simulate listener's emotion/impression of lyrics of music on a system, we have made 18 (9 emotion and 9 impression) models for 9 individual subjects. We have asked 9 individual subjects about the emotion/impression of 79 sample lyrics of music through a web questionnaire system. We used 79 Japanese pop songs in the RWC music library [9] as sample lyrics. The subjects do not listen to the music and only read the lyrics of music on the web page and answer emotion/impression of the lyrics. We have made a individual model from the individual's questionnaire data and the representative words of individual by using Adaptive Boosting with Logistic regression model on the Weka [10].

Table 4. The accuracy rates of models and the number of which subjects have felt impressions, as for Clear, Light and easy, Sensational and Calm (All sample music are 79).

	Clear		Light and easy		Sensational		Calm	
Subjects	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate
A	1	98.7	4	89.9	9	81.0	7	91.1
B	1	98.7	1	98.7	9	78.5	1	96.2
C	10	83.5	14	72.2	20	58.2	12	72.2
D	12	75.9	6	86.1	12	69.6	22	65.8
E	14	68.4	3	96.2	21	50.6	7	87.3
F	3	93.7	1	88.6	1	91.1	36	48.1
G	9	77.2	5	78.5	18	65.8	16	67.1
H	3	86.1	2	89.9	5	88.6	6	74.7
I	0	---	4	83.5	4	94.9	3	78.5
Baseline	13	67.1	9	79.8	20	58.2	29	50.6

Table 5. The accuracy rates of models and the number of which subjects have felt emotions, as for Joyful, Impressed, Touched and Enchanted (All sample music are 79).

	Dynamic		Swinging		Sentimental		Romantic	
Subjects	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate	Number	Accuracy Rate
A	5	92.4	20	64.6	33	44.3	13	55.7
B	15	60.8	16	65.8	33	40.5	8	79.7
C	21	49.4	24	58.2	25	58.2	35	49.4
D	22	59.5	11	78.5	45	60.8	31	55.7
E	12	68.4	11	78.5	37	41.7	31	46.8
F	24	55.7	21	55.7	8	83.5	15	67.1
G	22	63.3	19	70.9	18	55.7	17	57.0
H	11	81.0	4	94.9	38	54.4	16	77.2
I	14	73.4	25	50.6	45	63.3	12	64.6
Baseline	34	48.1	39	59.5	61	65.8	42	58.2

In order to evaluate individual model, we have made a baseline model by using the data that two subjects out of all subjects feel emotion/impression of lyrics. We have evaluated the estimate accuracy of individual model based on the 10-hold cross-validation.

3.1 Result and consideration concerning emotion model

Following tables show the number of which each subject has experienced emotions, and the rate of correctly classified data of the individual model.

Table 2 is the result concerning “Refreshed”, “Exhilarated”, and “Comfortable”. The result shows that the accuracy rate of them is over 70%. Individual models are more effective than the base model. The accuracy rates of individual models are better than the base model. However, many subjects do not feel these emotions from sample lyrics. Therefore, we think that the individual models only estimate their emotion concerning the limited representative words.

Table 3 is the result concerning “Joyful”, “Impressed”, “Touched” and “Enchanted”. All subjects have experienced these emotions from sample lyrics and have felt same emotion from same lyrics. However, the accuracy rate of individual model is over 60% and is not good. We could pick up a number of representative words that are related to these emotions. For example, as for the representative words related to “Joyful”, we have picked up the words that show human action (e.g. “run” and “sing”) and the words positive impression (e.g. “laugh” and “beautiful”). However, we also have picked up the words that afford negative impression to us (e.g. “afraid” and “wondering”). One reason of which we have picked up the word not related to the emotion is that there are many kinds of the representative word related to a word.

For example, there are words (e.g. “outstandingly”, “sweet voice” and “splendor”) as representative words of “beautiful” in a synonym dictionary. Therefore, we estimate the context of lyrics and have to select the representative word suitable for the word.

3.2 Result and consideration concerning impression model

Table 4 is the result concerning “Clear” and “Light and easy”. The result shows that almost same tendency of table 1. The accuracy rate of them is over 70% but many subjects do not feel these impressions from sample lyrics. As for “Clear”, we have picked up the representative words from a number of lyrics (e.g. “search”, “look at” and “believe”). As for “Light and easy”, we could not detect the good representative words. We think that one of reasons is that subjects have paid attention to except for verb and adjective. Table 4 also shows the result concerning “Sensational” and “Calm”. The number of which each subject experienced impression is different but we have detected a number of common representative words in subjects. We think that some subjects have same criteria about these impressions and feel same impressions from the lyrics.

Table 5 is the result concerning “Dynamic”, “Swinging”, “Sentimental” and “Romantic”. Subjects have felt these impressions from many sample lyrics. The representative words related to “Dynamic” are “move”, “dance”, “take a step” and “run”. The representative words related to “Swinging” are “laugh”, “shine” and “believe”

4 Selecting the representative words more related to emotion/impression

Our aim is to recommend the lyrics with the representative words as the reason of recommendation. In order that a user evaluates the candidate lyrics through the representative words, our system has to reduce the representative words to the number at which they can take a glance.

Table 6. The comparison of the accuracy rate between the normal baseline model and the reduction baseline model

	Emotion Impression	Normal baseline model (Using all representative words)	Reduction baseline model (Using the selected representative words)
Emotion	Joyful	48.1	65.8
	Comfortable	58.2	86.1
	Happy	62.0	83.5
	Impressed	43.0	79.7
	Serenity	44.3	72.2
	Touched	48.2	72.2
	Enchanted	50.6	79.7
Impression	Dynamic	48.1	70.9
	Sensational	58.2	78.5
	Swinging	59.5	74.7
	Sentimental	65.8	84.8
	Clam	50.6	72.2
	Romantic	58.2	54.4
	Spectacular	51.9	74.7

In order to reduce representative words with maintaining explanation ability, we have selected the representative words more related to emotion/impression of which the value of information gain is over 0.0, from all 4626 representative words. We have evaluated efficiency of selection of representative words by comparing the accuracy rate of the normal model and the reduction model. In this experiment, we evaluate only the baseline model of each emotion/impression and we do not evaluate the each subject's model. We also do not evaluate the baseline models based on data of which two subjects out of all subjects do not have experienced emotion/impression in over 20% of all sample lyrics. Therefore, we have evaluated the baseline models of seven emotions and the seven impressions, as shown in table 6.

As for emotion, for example, the selected representative words of "serenity" are "keep" and "think" and this model recommends the lyrics that included these words as serenity. The selected representative words of "touched" are "reach", "feel" and "gaze". We think that a user can be satisfied with the recommend reason concerning above representative words.

As for "Joyful", the authors think that "run", "laugh" and "sing" are relevant to joyful, but these words have not been chosen as the selected representative words by information gain.

As for impression, the accuracy rates of the six reduction models except for "romantic" are good than normal model. For example, the selected representative words of "dynamic" are "dance" and "encounter". The selected representative words of "calm" are "feel" and "keep". On the other hand, the accuracy rate of the reduction model of "romantic" is down than normal model. The selected representative words of "romantic" are "sweet", "reach" and "beautiful". The one of reasons is that we cannot have deleted the useless words (e.g. "go out" and "can do").

5 Conclusion

We have introduced the model of emotion/impression based on analyzing lyrics of music in order to recommend the music suitable for a user's emotion/impression. Our method explains the recommendation reason of music to a user through showing the words related to the emotion/ impression of the lyrics. We have selected the verb and adjective of lyrics in order to analyze the relationship between user's emotion/impression of lyrics. We have summarized the verb and adjective to the representative words. We have used the 4626 representative words of lyrics that are related to emotion/impression to make the models of 9 emotions and 9 impressions by using Adaptive boosting with logistic regression. We have gotten 9 subjects of 9 emotions and 9 impressions concerning 79 sample lyrics through a subjective evaluation experiment. We have made the models for 9 subjects and the baseline model and evaluate the models through 10-hold cross-validation.

The evaluations of the models have shown that the almost personalized models are better than the baseline model. The result shows that each subject has the personal relationship between his/her emotion/impression and the representative words in lyrics. We also have detected the representative words more related to emotion/impression by using information gain. We can use the representative words to explain the reason to a user.

In future works, we will try to improve the method of the summarizing of words in lyrics to the representative words by estimating the context of lyrics. We also will try to detect the representative words related to individual emotion / impression.

Acknowledge

This work was supported by JSPS KAKENHI Grant Number 25330331.

Reference

1. Lee, J. H. and Downie, J. S.: Survey of Music Information Needs, Uses, and Seek- ing Behaviours: Preliminary Findings. In: 5th International Conference on Music Information Retrieval, pp.441-446. (2004)
2. Yang, D. and Lee, W.:Music Emotion Identification from Lyrics. In: 11th IEEE International Symposium on Multimedia, pp.624-629, (2009)
3. Guo, Z., Wang, Q., Liu G., Guo, J. and Lu Y.: A Music Retrieval System Using Melody and Lyric. In: 2012 IEEE International Conference on Multimedia and Expo Workshops, pp.343-348. (2012)
4. Hu, Y., Chen, X. & Yang, D. : Lyric-Based Song Emotion Detection with Affective Lexicon and Fuzzy Clustering Method. In: Proceedings of the 10th International Society for Music Information Retrieval Conference, pp. 123-128. (2009).
5. Laurier, C., Crivolla, J., and Herrera P.: Multimodal Music Mood Classification Using Audio and Lyrics, In: Proceedings of ICMLA 2008, pp. 688–693. (2008).
6. Wang, T., Kim, D., Hong K. and Youn, J.: Music Information Retrieval System using Lyrics and Melody Information, In: 2009 Asia-Pacific Conference on Information Processing, pp.601-604.(2009).

7. Seungwon, O., Minsoo, H. and Jinsul K.: Music Mood Classification Using Intro and Refrain Parts of Lyrics, In: 2013 International Conference on Information Science and Applications, pp.1-3. (2013).
8. Chung-Yi, C. Ying-Shian, W. Wei-rong, C., Wu, D.C., Hsu, J.Y.-j. and Tsai, R.T.-H.,The power of words: Enhancing music mood estimation with textual input of lyrics, In: 3rd International Conference on Affective Computing and Intelligent Interaction (ACII 2009), pp.1-6.(2009).
9. Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, In: Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002), pp.287-288, (2002).
10. Weka, http://www.cs.waikato.ac.nz/ml/weka/
11. MeCab, http://taku910.github.io/mecab/