Non Keyword-Based Music Retrieval Using Social Tags

Chang Bae Moon
ICT-Convergence Research Center
Kumoh National Institute of Technology
Gumi, South Korea
cb.moon@kumoh.ac.kr

Jong Yeol Lee
Department of Computer Software Engineering
Kumoh National Institute of Technology
Gumi, South Korea
soyeum@kumoh.ac.kr

Dong-Seong Kim
Department of IT Convergence Engineering
Kumoh National Institute of Technology
Gumi, South Korea
dskim@kumoh.ac.kr

Byeong Man Kim
Department of Computer Software Engineering
Kumoh National Institute of Technology
Gumi, South Korea
bmkim@kumoh.ac.kr

***Abstract*—** To retrieve music by mood tags in a social network, we introduce a mood vector, which allows moods of music pieces and mood tags to be represented Pinternally by numeric values. A mood vector consists of 12 arousal-valence pairs each represents a mood of Thayer's two-dimensional mood model. To determine the mood vector of a music piece, a Support Vector regressor is created for arousal and valence each using features of a music piece. Then, the regressors predict a mood vector. To map a mood tags to its mood vector, we investigate the relationship between them based on tagging data retrieved from Last.fm. To show the benefits of the proposed method, in this paper, we create a test set by using last.fm tags and their synonyms, and measure its retrieval performance over the keyword-based approach using the test set. The results illustrate that the proposed method can be useful in many respects including solving the problem caused by synonyms.

Keywords—Music mood social tag; Mood vector; Relationship between mood and tag

I. INTRODUCTION

For social mood tags, that is, mood tags in social networks to be used as queries of music retrieval systems, some problems should be addressed. The first problem is that different words can be used identical meanings by different users; for example, "relaxed" and "calm" are different words, but a soothing piece of music may be tagged the different two words by two different users. The second problem is a tagging level; i.e., in the case of "placid" and "very placid", the same root word is used, but those words are expressed by different degrees.

In this paper, a music retrieval method by social mood tags is proposed, and it is shown that the method is useful in solving the synonyms problems. To this end, we introduced the mood vector (an arousal-valence pair representing a mood in Thayer's two-dimensional mood model) as an internal tag. Using this method, moods of music pieces and mood tags are all represented internally by numeric values; pieces having moods similar to the mood tags of a query can then be retrieved based on the similarity of their mood vectors, even if their tags do not exactly match the query.

II. RELATED STUDIES

Existing emotion models include the Russell model [1], the Hevner Model [2], and the Thayer model [3]. Since both the Russell and Hevner models use adjectives to describe emotions, ambiguity arises if adjectives have multiple meanings. For this reason, we used Thayer's two-dimensional model, in which each mood or emotion is expressed by two values, arousal and valence. Arousal refers to the strength of stimulation that listeners feel (i.e., weak or powerful) and valence refers to the intrinsic attractiveness (positive valence) or averseness (negative valence).

A folksonomy is a classification system in which volunteers collaboratively create and manage tags to annotate and categorize content. Also, the folksonomies can solve the expanding problems of taxonomies. It means that category of folksonomy can be expanded by volunteers without web manager. So, tags in a folksonomy are ones of social tags.

A number of studies have explored music folksonomy tags [4, 5, 6, 13]. In Laurier et al.'s and Kim et al.'s studies [6, 7], music mood tags from the well-known folksonomy site Last.fm were treated as categories, upon which these authors constructed classification models. These classification models first determine the category of each piece of music, and then the folksonomy tag corresponding to the category is applied. In Steven et al.'s study [4], music was subdivided into sub-units and features were extracted. Then, these features were learned using a Support Vector Machine (SVM). When a new music piece is inputted, a mood tag is assigned based on the classification model.

III. MUSIC MOOD RETRIEVAL USING SOCIAL TAGS

The music retrieval system that we consider consists of four phases. The first phase is to create a prediction model for each mood. The second phase is to obtain the mood vector of the music piece and attach it as its internal tag. The third phase is to define the mapping relationship between mood tags and their mood vectors. The last phase is to retrieve music using mood vectors and tag-mood mapping information.

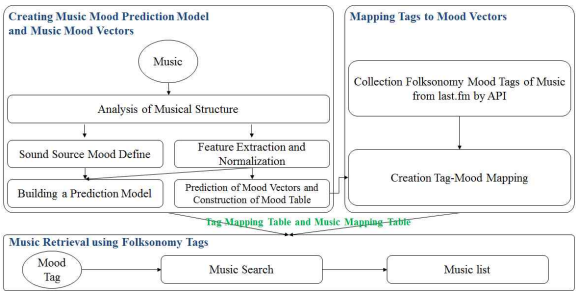

Fig. 1. Overview of music retrieval system

A. Creating Music Mood Vectors and Music Mood Prediction Model

Last.fm, a prominent online example of music folksonomy in action, boasts more than 1,000 music pieces that have at least one music mood tag. These pieces can be collected using API (Application Programming Interface). However, users provide mood tags in the form of words, not in the form of a mood vector. In order to accurately translate these tags and reflect users' individual intended meanings for terms with multiple possible meanings, we would need to obtain individual mood vectors from all users. However, such an approach is impractical, and would be both time-consuming and prohibitively expensive. For this reason, we built models to predict the mood vector of a given music piece using existing music mood data [8].

The prediction models are built through the three steps as show in Figure 1. In the analysis step of musical structure, music pieces are separated into segments through musical structural analysis [8, 9, 10]. Then, three music segments are chosen for each piece by moon et al. [8]: one from the "Intro" section, one from the "Outro" section, and the one with the highest energy called the Representation section.

In the feature extraction and normalization step, we use the 391 features extracted from Lartillot's MIR toolbox [7]. However, when features are extracted using MIR toolbox, NaNs - values that cannot be expressed numerically - may occur. So, features with at least one NaN are removed. Thus, 330 features are used in our experiments and feature values are normalized between –1 and +1.

In the building step, we use Support Vector Regression (SVR) to build regressors. The SVR is provided by Support Vector Machines (LIBSVM) [9, 11, 12], which has recently become more common for regression analysis. The input value of the regressors is the normalized feature vector of a music segment, which is extracted as described above. The data from Moon et al.'s study [8] are used to create predictors of mood vectors (See Figure 2).

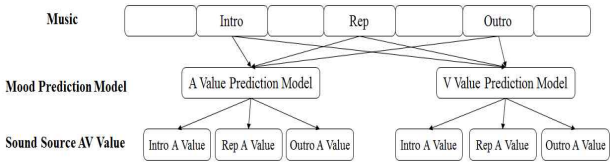

Fig. 2. Mood Vectors(AV Value) Prediction

In predicting mood vectors of music pieces step, we use the 1,243 music pieces on Last.fm with at least one mood tag as our sample and predict their mood vectors. Using the method described in "Analysis of musical structure", three segments are selected per piece; a music vector is then created for each segment (2 or 3 per piece). The mood vector of each segment is predicted using the mood vector predictor depicted, which is henceforth referred to as a mood table of music, music–mood table, or music–mood mapping table.

B. Mapping Tags to Mood Vectors

The method used to retrieve music using mood tags is as follows: first, the user inputs the tag to be retrieved (query in Figure 3); the mood vector of the query tag is then searched in the tag–mood mapping table; then, music pieces with mood vectors similar to the mood vector of the query tag are retrieved.

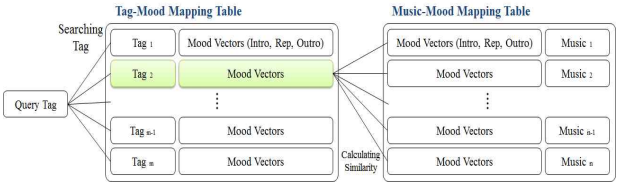

Fig. 3. Music Retrieval using Folksonomy Tags

IV. Experiments and Analysis

Although all mood words could be used as query tags to measure the retrieval performance of the suggested method, only the 12 words from Thayer's two-dimensional mood model are considered due to the limitations on processing such a vast data set. Synonyms (provided by www.synonym.com) are used to build the answer set for the 12 mood words; For example, the tag "peaceable" is grouped with "peaceful," as "peaceable" is a synonym of the basic mood adjective "peaceful." Each music piece has at least one mood tag; the number of music pieces associated with each mood are calm (501), pleased (241), sad (527), excited (304), nervous (151), peaceful (90), relaxed (527), happy (290), bored (190), sleepy (79), angry (205), and annoying (242).

We compare the retrieval performance of the proposed method with the traditional retrieval method, the keyword-based method from the two respects. The first is to compare the retrieval performance of the 12 mood words mentioned above as queries and the second is to compare the retrieval performance in the case that two words are included in a tag.

The retrieval performance of the keyword-based method for 12 mood words is in [Table 1] and the recall rate of [Table 1] is that of the case where synonyms are considered. Further, the precision is 1.0, which is due to the fact that only music pieces having same text with a query tag are retrieved. Checking this in details, tag 'Peaceful' shows recall rate of 0.74, which means 74% of the 90 music pieces having "Peaceful" tag or its synonym tag are retrieved. Tag 'Pleased' shows the lowest recall rate of 0.01, because only 3 pieces of music include tag 'Pleased' among 241 relevant music pieces.

TABLE I. RETRIEVAL PERFORMANCE OF THE KEYWORD-BASED METHOD (NOMPIS : NUMBER OF MUSIC PIECES INCLUDING SYNONYM, NOMPNIS : NUMBER OF MUSIC PIECES NOT INCLUDING SYNONYM)

Tag	NoMPiS	NoMPniS	Recall rate
Calm	501	171	0.34
Pleased	241	3	0.01
Sad	527	303	0.57
Excited	304	38	0.13
Nervous	151	33	0.22
Peaceful	90	67	0.74
Relaxed	527	102	0.19
Happy	290	208	0.72
Bored	190	37	0.19
Sleepy	79	56	0.71
Angry	205	75	0.37
Anooying	242	26	0.11
Average Recall Rate			**0.36**

The proposed method, however, shows the precision of 0.64 (average precision of experiments repeated in 10 times) at recall level 0.1 and the precision of 0.33 at recall level 0.2. Namely, although the precision of the proposed method may be lower than that of the keyword-based method, more music can be provided for some queries, for example 'Pleased', because the proposed method considers not simple text but the contents of music and thus music with similar moods can be provided to users. In [Table 2], Intro means the mood vector of the Intro section of a music piece is used. Similarly, Outro means the Outro section, Representation means the Representation section and All means the three sections all.

TABLE II. RETRIEVAL PERFORMANCE OF THE PROPOSED METHOD FOR 12 MOOD WORDS

Recall level	Intro	Represent ation	Outro	All
0.1	0.50	**0.64**	0.51	0.58
0.2	0.30	0.33	0.31	0.32
0.3	0.27	0.31	0.29	0.30
0.4	0.26	0.29	0.28	0.29
0.5	0.26	0.28	0.27	0.28
0.6	0.25	0.27	0.27	0.27
0.7	0.25	0.26	0.26	0.27
0.8	0.25	0.26	0.26	0.26
0.9	0.24	0.25	0.25	0.25
1	0.23	0.24	0.24	0.24

As shown in [Table 3], the retrieval performance of two tags, 'happy songs' and 'sad songs' where each tag includes two words, is given. The precision of 0.81 & 0.94 is seen for each tag at recall level 0.1. Although the precision of the keyword-based method is 1.0, the recall rate of tag 'happy songs' is 0.04(11/290) when considering synonym and recall rate is 0.09(49/527) for tag 'sad songs'. Notice that only 11 music pieces have tag 'happy songs' and only 49 pieces 'sad songs'. In consequence, although the keyword-based method shows better performance than the proposed method in precision, the keyword-based method provides only music pieces having tags matched with the query tag exactly while the proposed method provides music pieces having different tags but similar mood.

TABLE III. RETRIEVAL PERFORMANCE WHEN A QUERY INCLUDES TWO WORDS

Recall Level	Tag 'happy songs'		Tag 'sad songs'	
	Proposed	Keyword-Based	Proposed	Keyword-Based
0.1	0.81	1.00 (11/290)	0.94	1.00 (49/527)
0.2	0.38		0.59	
0.3	0.35		0.59	
0.4	0.32		0.58	
0.5	0.29		0.56	
0.6	0.27		0.52	
0.7	0.27		0.50	
0.8	0.26		0.46	
0.9	0.25		0.45	
1	0.23		0.43	

V. CONCLUSION

In this paper, to solve the synonym problem associated with social tags, the mood vector of music was introduced as an internal tag. A mood vector consists of two values indicating the arousal and valance of a music piece. Music pieces were tagged internally using these numeric values, enabling the retrieval of music with similar moods. To implement a retrieval system based on internal tags, music vectors should be generated for both music pieces and folksonomy tags.

The paper focused on not the retrieval performance but the benefits of the internal numeric tagging of music over the keyword-based approach. So, we need to focus on the retrieval performance at the next work. The retrieval performance of the approach could be improved by enhancing the predictive power of mood vectors, which is largely dependent on the quality and quantity of the training data provided.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03033733, 2018R1C1B6001042).

REFERENCES

[1] Russell, J.A., circumplex model of affect. Journal of Personality and Social Psychology, No. 39, 1161. 1980.

[2] Hevner, K., xperimental studies of the elements of expression im music. The American Journal of Psychology, Vol. 48, No. 2, 246–68. 1936.

[3] Thayer, R.E., The Biopsychology of Mood and Arousal, New York. Oxford University Press., 1989.

[4] Steven R. Ness, Anthony Theocharis, George Tzanetakis and Luis Gustavo Martins, Improving Automatic Music Tag Annotation Using Stacked Generalization Of Probabilistic SVM Outputs. Proc. of ACM MM'09, pp.705-708, 2009.

[5] Laurier, C., Meyers, O., Serra`, J., Blech, M., Herrera, P., Music Mood Representation from Social Tags. Pro- ceedings of the 10th International Society for Music Information Conference, Kobe, Japan, 2009.

[6] Kim, J.H., Lee, S., Kim, S.M., Yoo, W.Y., Music mood classification model based on Arousal- Valence values. In: Proc. ICACT. pp. 292–295, 2011.

[7] Lartillot, O. and Toiviainen, P., A Matlab toolbox for musical feature extraction from audio. Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), pp. 237-244, Bordeaux, France, September 10-15, 2007.

[8] Moon, C.B., Kim, H.S., Lee, H.A. and Kim, B.M., Analysis of relationships between mood and color for different musical preferences, Color Research & Application, vol. 39, Issue 4, pp. 413-423, 2014.

[9] Lee, J.I., Yeo, D.-G., kim, B.M., Lee, H.-Y., Automatic Music Mood Detection through Musical Structure Analysis. International Conference on Computer Science and its Application CSA 2009, pp. 510-515, 2009.

[10] Levy, M. Sandler, M. and Casey, M., Extraction of High-Level Musical Structure From Audio Data and Its Application to Thumbnail Generation. Proc. of ICASSP'06, Vol. 5, Toulouse, France, 13-16, 2006.

[11] Ryu, S.-J., Lee, H.-Y., Cho, I.-W. and Lee, H.-K., Document Forgery Detection with SVM Classifier and Image Quality Measure. Lecture Notes in Computer Science, vol.5353, pp.486-495, 2008.

[12] Chang, C.-C. and Lin, C.-J., LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/ ~ cjlin/ libsvm/, 2001.

[13] Mark Levy and Mark Sandler, Music Information Retrieval Using Social Tags and Audio, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 11, NO. 3, pp. 383-394, APRIL 2009.