The 4th ICIBA 2015, International Conference on
Information Technology and Engineering Application
Palembang-Indonesia, 20-21 February 2015

Music Retrieval using Mood Tags in Folksonomy

Chang Bae Moon, HyunSoo Kim, Dong-Seong Kim, Sung-Phil Heo, Byeong Man Kim

Kumoh National Institute of Technology,Gumi, Gyeongbuk, South Korea
e-mail: cb.moon@kumoh.ac.kr

Abstract

This paper propose a music retrieval using mood tats in folksonomy for solving Synonym problems. Normally, a music piece and a mood tag can be represented by mood numeric vectors internally. To determine the mood vectors of a music piece, 12 regressors are created by Support Vector Regression using features of a music piece. Then, a mood vector is predicted by the 12 regressors. To map a folksonomy mood tags to its mood numeric vectors, the relationship between mood vectors of music piece and the folksonomy mood tags was investigated based on tagging data retrieved from Last.fm. To evaluate retrieval performance, music pieces on Last.fm annotated with at least one mood tag were used as a test set. When calculating precision and recall, music pieces annotated with synonyms of a given query tag were treated as relevant. These experiments on a real-world data set illustrate the utility of the internal tagging of music. Our approach offers a practical solution to the problem caused by synonyms.

Keywords : *Music mood, Folksonomy mood tag, Last.fm, Relationship between mood and tag*

1 INTRODUCTION

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, category of folksonomy can be expanded by volunteers without web manager.

However, folksonomies do tend to have problem relating to tags. 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 mood tags in Folksonomy is proposed, and the method is to solve the synonyms problems when retrieving music. To do this, we introduced the mood vector (12 values representing different moods according to Thayers 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 identified as 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.

2 RELATED STUDIES

Existing emotion models include the Russell model (Russell, J.A. (1980)), the Hevner Model (Hevner, K. (1936)), and the Thayer model (Thayer, R.E. (1989)). Since both the Russell and Hevner models use adjectives to describe emotions, ambiguity arises if adjectives have multiple meanings. For this reason, we used Thayers 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 aversiveness (negative valence).

A number of studies have explored music folksonomy tags (Steven R. Ness, et al. (2009); Laurier, C., et al. (2009); Kim, J.H., et al. (2011)). In Laurier et al.s and Kim et al.s studies (Laurier, C., et al. (2009); Kim, J.H., et al. (2011)), 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 (Steven R. Ness, et al.), 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.

3 MUSIC MOOD RETRIEVAL BASED ON FOLKSONOMY TAGS

The object of the present paper is composed of four phases (Fig.1). The first and second phases are to create a prediction model for each mood; once the mood vectors of the music piece are obtained, they are attached as internal tags. The third phase is to define the mapping relationship between folksonomy tags on Last.fm and their mood vectors. The last phase is to retrieve music using mood vectors and tag-mood mapping information.

3.1 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 (Moon, C.B., et al.(2014)).

The prediction models are built through the three steps as show in Figure 1. In the analysis step of musical structure, music pieces were separated into segments through musical structural analysis (Lee, J.I., et al.(2009); Moon, C.B., et al.(2014); Levy, M., et al.(2006)). Then, three music segments were chosen for each piece by moon et al. (Moon, C.B., et

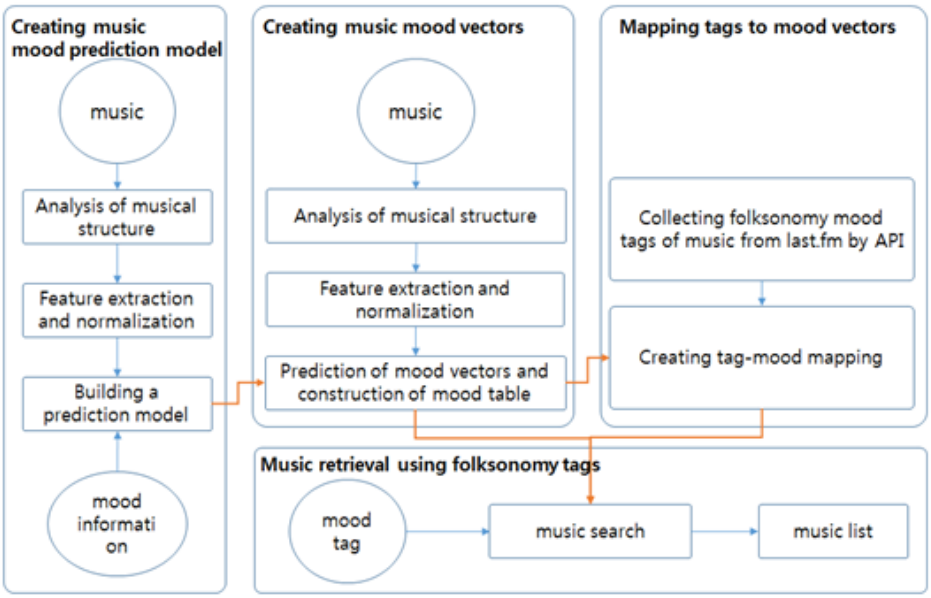


Figure 1: Overview of music retrieval system

al.(2014)): one from the Intro section, one from the Outro section, and the one with the highest energy.

In the feature extraction and normalization step, we used the 391 features extracted from Lartillots MIR toolbox (Lartillot, O., et al.(2007)). However, when features are extracted using MIR toolbox, NaNsvalues that cannot be expressed numericallymay occur. So, features with at least one NaN were removed. Thus, 330 features are used in our experiments and feature values were 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) (Lee, J.I., et al.(2009); Ryu, S.-J., , et al.(2008); Chang, C.-C., et al.(2001)), 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 als study(Moon, C.B., et al.(2014)) were used to create predictors of mood vectors.

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

3.2 Mapping tags to mood vectors

To obtain the mapping information for mood tags and mood vectors, we needed to collect mood tag data for all of the music pieces used in our analysis. On Last.fm, listeners can check the relevant tags of a music piece or retrieve music pieces by selecting a tag. The associated

music tags of a music piece are obtained automatically through API which requires the name of the artist and the track.

Different numbers of music pieces were associated with each tag. Therefore, we needed to define the representative mood vector of each tag. We did this by averaging the mood vectors for music pieces with the same tag. These averaged mood vectors are the mood vectors of a tag which were used to connect mood tags and music together with the musicmood table. Only 50% 70% of tag data were used to create maps; the remaining portion was used for performance evaluation.

3.3 Music retrieval using mood tags

The method used to retrieve music using mood tags is as follows: first, the user inputs the tag to be retrieved(query in Fig. 2); the mood vector of the query tag is then searched in the tagmood mapping table; then, music pieces with mood vectors similar to the mood vector of the query tag are retrieved.

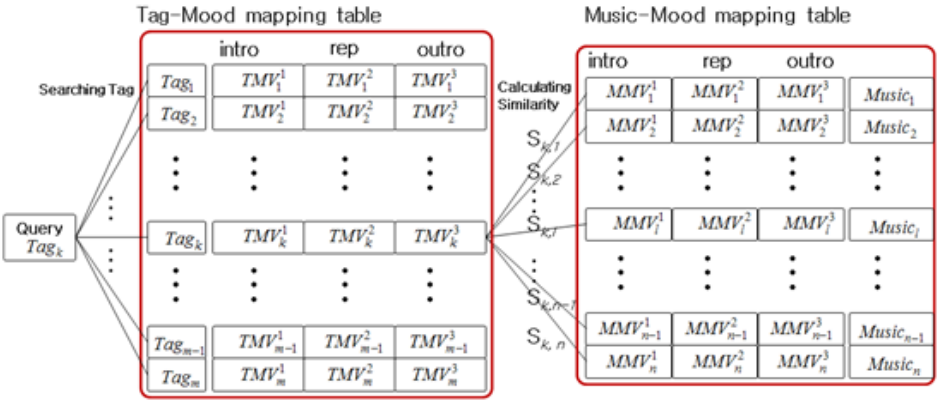


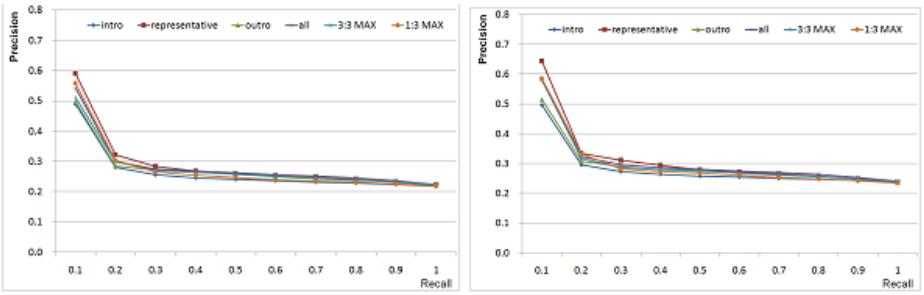
Figure 2: Music retrieval using folksonomy tags

4 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 Thayers two-dimensional mood model were considered due to the limitations on processing such a vast data set. Synonyms (provided by www.synonym.com) were used to build the answer set for the 12 mood words; For example, the tag peaceable was grouped with peaceful, as peaceable is a synonym of the basic mood adjective peaceful." Each music piece had 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).

The retrieval performance for the six combinations of mood vectors is given in Fig. 3 (a). The results come from using 50% of the music pieces to build the tag-mood table, and using the other 50% of the music pieces for testing. The best performance at recall level 0.1 was 0.49 for Intro, 0.59 for Representative, 0.50 for Outro, 0.54 for All, 0.51 for 3:3 MAX, and

0.56 for 1:3 MAX; Representative performed the best at most recall levels. When we changed the ratio of training: testing data from 50:50 to 70:30, the best performance was improved by 0.05 (from 0.59 to 0.64) (Fig. 3 (b)). It means, retrieval performance of our proposed method can be improved, when there are more training or testing data set.



(a) Ratio of training: testing data = 50:50 (b) Ratio of training: testing data = 70:30

Figure 3: Retrieval performance

5 CONCLUSION

In this paper, to solve the synonym problem associated with folksonomies, the mood vector of music was introduced as an internal tag. A mood vector consists of 12 values indicating the presence of the 12 moods from Thayers two-dimensional mood model. 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 demonstrated the internal tagging of music to be useful when combined with our scheme for solving the problem caused by synonyms. However, the retrieval performance of this 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.

6 ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center)) / CITRC(Convergence Information Technology Research Center) support program (NIPA-2014-H7501-14-1002) supervised by the NIPA(National IT Industry Promotion Agency)

References

Chang, C.-C., Lin, C.-J. (2001), LIBSVM: a library for support vector machines. Software available at *http://www.csie.ntu.edu.tw/cjlin/ libsvm/*.

Hevner, K. (1936), xperimental studies of the elements of expression im music. T*he American Journal of Psychology*, Vol. *48*, No. 2, 24668.

Kim, J.H., Lee, S., Kim, S.M., Yoo, W.Y.(2011), Music mood classification model based on Arousal- Valence values. *In: Proc. ICACT.* pp. 292295.

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

Lartillot, O. and Toiviainen, P.(2007), 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.

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

Levy, M. Sandler, M. and Casey, M. (2006), 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.

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

Russell, J.A. (1980), circumplex model of affect. J*ournal of Personality and Social Psychology*, No. 39, 1161.

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

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

Thayer, R.E. (1989), *The Biopsychology of Mood and Arousal*, New York. Oxford University Press.