2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)

Popular Music Estimation Based on Topic Model Using Time Information and Audio Features

Shohei Kinoshita
School of Engineering
Hokkaido University
N-13, W-8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan
Email: kinoshita@lmd.ist.hokudai.ac.jp

Takahiro Ogawa and Miki Haseyama
Graduate School of Information Science and Technology
Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: ogawa@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

***Abstract*—This paper presents popular music estimation based on a topic model using time information and audio features. The proposed method calculates latent topic distribution using Latent Dirichlet Allocation to obtain more accurate music features. In this approach, we also use release date information of each music as time information for concerning the relationship between music trends and each age. Then, by using the obtained latent topic distribution features, the estimation of the popular music becomes feasible based on a Support Vector Machine classifier. Experimental results show the effectiveness of our method.**

I. Introduction

With the growth of online music services such as iTunes Store[1], users can easily search a large amount of music data to obtain their desired ones. Therefore, many researchers have proposed music recommendation methods based on content-based filtering using audio signals of music data and their metadata [1,2]. Note that in order to improve the performance of the music recommendation, it is important to grasp music trends in the past and in the future. In this paper, we focus on popular music estimation which is one of the simplest problems for finding music trends.

It is expected that popular music estimation can be realized by generating classifiers which judge whether each music is one of popular music or not. In this approach, we have to consider two important points. First, features used for the classification should be able to represent music trends. Second, since music trends change according to each age, their relationship should be concerned.

In this paper, we propose a popular music estimation method based on a topic model, Latent Dirichlet Allocation (LDA) model [3], using both time information and audio features. In the same manner as our previous work [4], we can calculate Audio Word (AW) features from audio signals of music to obtain their latent topics. In this approach, we also use their release date as a feature for obtaining AW features which can concern the relationship between music trends and each age. Then, by defining the obtained latent topic distribution of each music as a new feature vector, we can perform the popular music estimation with a Support Vector Machine (SVM) classifier [5] generated from popular and non–popular music.

[1]http://www.apple.com/jp/itunes/

II. Popular Music Estimation

In this section, we present the proposed popular music estimation method. The proposed method consists of three procedures, AW feature calculation using audio signals and release date of music, LDA-based calculation of latent topic distribution and SVM-based estimation of popular music. Therefore, these three procedures are respectively explained in II–A, II–B and II–C.

A. AW Feature Calculation

In order to apply LDA to music, we calculate the AW features as shown in our previous work [4]. First, the proposed method divides each music m_i ($i = 1, 2, \cdots, N$; N being the number of music) into several frames f_i^j ($j = 1, 2, \cdots, M_i$; M_i being the number of frames in music m_i) in the same intervals and extracts their audio features shown in Table I. Furthermore, by combining the obtained 151 audio features and release date of each music m_i, a 152–dimensional feature vector $\boldsymbol{v}_i^j$ ($j = 1, 2, \cdots, M_i$) is obtained for each frame. By using the release date, we can concern the time variation of popular music trends. From all feature vectors $\boldsymbol{v}_i^j$, we perform *k*-means clustering [6] to obtain AWs corresponding to the cluster centers. Then, by applying a Bag-of-Features method [7] to each music, we can obtain its AW feature vector $\boldsymbol{w}_i$ ($i = 1, 2, \cdots, N$).

B. LDA-based Calculation of Latent Topic Distribution

In the proposed method, we generate the LDA model from AW feature vectors $\boldsymbol{w}_i$ ($i = 1, 2, \cdots, N$). Then each music can be represented by mixing ratio of some latent topics. Therefore, we can define the latent topic distribution of each music as a new feature vector $\boldsymbol{l}_i$ ($i = 1, 2, \cdots, N$). Note that when applying the LDA model to music data, we exclude the AWs which have the highest and lowest 5% appearance frequency as the Stop Words (SWs) [8]. The SWs are the AWs which appear almost all music or only few music. Then we can obtain the high-precision topic model by excluding the SWs.

C. SVM-based Estimation of Popular Music

From the feature vectors $\boldsymbol{l}_i$ ($i = 1, 2, \cdots, N$) obtained by LDA, we calculate a classifier based on the SVM. Note that we assume that we have previously known whether each music m_i is popular music or not. Then given a new music m whose release date is only known, we calculate its feature vector $\boldsymbol{l}$

TABLE I. FEATURES USED FOR AUDIO SIGNALS IN THE PROPOSED METHOD.

CATEGORY	DESCRIPTION	STATISTICS	DIMENSION
dynamics	Root Means Square	Mean, Std	2
spectral	Centroid	Mean, Std	2
	Brightness	Mean, Std	2
	Spread	Mean, Std	2
	Skewness	Mean, Std	2
	Kurtosis	Mean, Std	2
	Rolloff	Mean, Std	4
	Entropy	Mean, Std	2
	Flatness	Mean, Std	2
	Roughness	Mean, Std	2
	Irregularity	Mean, Std	2
timbre	Zero Crossing Rate	Mean, Std	2
	MFCC	Mean, Std	26
	Low Energy	Mean, Std	2
tonal	Key Strength	Mean, Std	48
	Chromagram	Mean, Std	24
	Key	Mean, Std	2
	Tonal Centroid	Mean, Std	12
	Mode	Mean, Std	2
rhythm	Tempo	Mean	1
	Pulse Clarity	Mean, Std	2
	Event Density	Mean, Std	2
	Attack Time	Mean, Std	2
	Attack Slope	Mean, Std	2
TOTAL			151

in the same manner as shown in the previous subsection and perform the classification using the obtained classifier. From the obtained result, we can estimate whether the new music m is popular music or not.

III. EXPERIMENTAL RESULTS

In this section, we show experimental results to verify the performance of the proposed method. In this experiment, we used 220 music (100 music are popular ones and the others are not). The popular music are included in the top five music of the single CD sales charts of each year from 1994 to 2007 in Japan, and Billboard JAPAN Hot 100 Year End[2] of each year from 2008 to 2013. The non–popular music are not in the top 100 music of the single CD and the album CD sales charts of each year from 1994 to 2013 in Japan. First, we calculated the features proposed in Section II from the beginning 120 seconds of each music. When we divided each music into several frames, the frame length was set to 0.5 second, and its slide length was set to 0.25 second. Next, we converted the release date 1994, 1995, ..., 2013 to values 5, 10, ..., 100 and used these values as features. When AWs are calculated by k–means clustering, the number of the clusters was empirically set in the range of 1000, 1500, ..., and 5000. Furthermore, the number of latent topics was set in the range of 10, 15, ..., and 50 in the LDA model. In our experiment, we adopted leave–one–out evaluation. For evaluating the performance of our method, we use Recall, Precision, F-measure and Accuracy defined as follows:

$$\text{Recall} = \frac{N_a}{N_b}, \tag{1}$$

$$\text{Precision} = \frac{N_a}{N_c}, \tag{2}$$

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall + Precision}}, \tag{3}$$

$$\text{Accuracy} = \frac{N_d}{\text{Number of all music}}, \tag{4}$$

[2]The hit charts based on sales, airplay and iTunes. http://www.billboard-japan.com/

TABLE II. RESULTS OF THE PROPOSED METHOD AND THE COMPARATIVE METHODS.

	Recall	Precision	F-measure	Accuracy
Proposed method	**0.83**	**0.790**	**0.810**	**0.823**
Comparative method 1	0.76	0.768	0.764	0.786
Comparative method 2	0.67	0.770	0.717	0.759

where N_a is the number of music estimated accurately as "popular music", N_b is the number of "popular music", N_c is the number of music estimated as "popular music", and N_d is the number of music estimated accurately as "popular music" or "non-popular music".

In this experiment, we employed two comparative methods: a method which does not use the release date information (comparative method 1), a method using only $\boldsymbol{w}_i$ without generating the LDA model (comparative method 2).

The results are shown in Table II. Note that we set the number of clusters and the latent topics to the values which output the best accuracy. As shown in this table, all evaluation criteria of the proposed method outperform those of the comparative methods. Comparing results of our method and comparative method 1, we can confirm the effectiveness of introducing the release date information. Furthermore, we can confirm the effectiveness of using the LDA model by comparing the results of our method and comparative method 2. Hence, we can verify the effectiveness of our method.

IV. CONCLUSION

In this paper, we have proposed a method for estimating popular music based on LDA using audio features and time information. By using time information, we can concern music trends of each age. Furthermore, by using the topic model, i.e., the LDA model, we can successfully represent features of each music. Experimental results are shown to verify the effectiveness of our method.

V. ACKNOWLEDGMENT

This work was partly supported by Grant-in-Aid for Scientific Research (B) 25280036 from JSPS.

REFERENCES

[1] B. Loga, "Music recommendation from song sets," in *Proc. The International Society for Music Information Retrieval Confernce*, 2004, pp. 425-428.

[2] K. Hoashi, K. Matsumoto, and N. Inoue, "Personalization of user profiles for content-based music retrieval based on relevance feedback," in *Proc. ACM International Conference on Multimedia*, 2003, pp. 110-119.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol.3, pp. 993-1022, 2003.

[4] R. Nakanishi, T. Ogawa, M. Haseyama, "A note on probabilistic topic modeling with user preferences and its application to music recommendation (in Japanese)," *The Institute of Image Information and Television Engineers*, 2014, vol. 38, no. 7, pp. 313-318.

[5] C. Cort and V. Vapnik, "Support-vector networks," *Machine Learning*, vol.20, pp.273-297, 1995.

[6] P.S. Bradley and U.M. Fayyad, "Refining initial points for k-means clustering," in *Proc. 15th International Conference on Machine Learning*, 1998, pp. 91-99.

[7] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray "Visual categorization with bags of keypoints," in *ECCV: Workshop on Statistical Learning in Computer Vision*, pp.59-74, 2004.

[8] J. Sivic and A. Zisserman, " Video Google: a text retrieval approach to object matching in videos," in *Proc. 9th IEEE International Conference on Computer Vision*, 2003, vol.2, pp.1470-1477.