LIA @ MediaEval 2013 MusiClef Task: A Combined Thematic and Acoustic Approach

Mohamed Morchid, Richard Dufour, Mohamed Bouallegue,
Georges Linarès and Driss Matrouf*
LIA - University of Avignon
Avignon, France
{firstname.lastname}@univ-avignon.fr

ABSTRACT

In this paper, we describe the LIA system proposed for the MediaEval 2013 Soundtrack task. The aim is to predict the most suitable soundtrack from a list of candidate songs, given a TV commercial. The organizers provide a development dataset including multimedia features. The initial assumption of the proposed system is that commercials which sell the same type of product, also share the same music rhythm. A two-fold system is proposed to provide a music for a commercial: find commercials with close subjects in order to determine the mean rhythm of this subset, and then extract from the candidate songs the music which better correspond to this mean rhythm.

1. INTRODUCTION

The success of a product or a service essentially depends of the way to present it. Thus, companies pay much attention to choose the most appropriate advertisement that will make a difference in the customer choice. The advertisers have different media possibilities, such as journal paper, radio, TV or Internet. In this context, they can exploit the audio media using a song related to the commercial which attracts listeners. Therefore, the choice of an appropriate song is crucial and can determine the success of a product [5, 2].

For these reasons, the MediaEval 2013 Soundtrack task for commercials becomes a challenging and helpful task [3]. Indeed, the MusiClef task seeks to make this process automated by taking into account both context- and content-based information about the video, the brand, and the music. The main difficulty of this task is to find the set of relevant features that best describes the most appropriate song for a video. We propose a hybrid approach that uses a set of features from textual and audio media.

2. PROPOSED APPROACH

The proposed hybrid system is composed of two processes. The first one projects a TV commercial into a topic space to find a set of other commercials sharing close topics. A TV commercial from the test set is thus linked to the TV commercial from the development set sharing the closest topics. As a result, each TV commercial from the test set will be associated with a song extracted from the development data.

The second step has the responsibility to find, using audio features, the most similar songs to the one associated during the first step from a list of candidate songs (see figure 1).

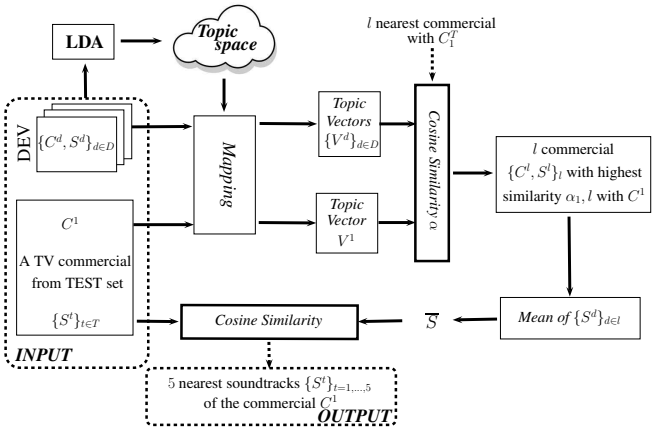

Figure 1: Global architecture of the proposed system.

In details, the development set D is composed of TV commercials C^d, with for each, a soundtrack S^d and a vector representation V^d related to the d^{th} TV commercial. In the same manner, the test set T is composed of TV commercials C^t, with, for the t^{th} one, a vector representation V^t and a soundtrack S^t to predict. Then a similarity score $\{\alpha_{d,t}\}_{d=1,\ldots,D}^{t=1,\ldots,T}$ is computed for each commercial C_i^d of the development set given one from the test set C^t:

$$D = \{C^d, V^D, S^d\}_{d=1,\ldots,D} \qquad (1)$$
$$T = \{C^t, V^T, S_k^t\}_{t=1,\ldots,T}^{k=1,\ldots,5000} .$$

In the next sections, the topic space representation and the mapping of a commercial in this topic representation are described. Then, the computed similarity score is detailed. Finally, the soundtrack prediction process from a TV commercial is explained.

2.1 Topic representation of a TV Commercial

Let's consider a corpus D from the development set of TV commercials with a word vocabulary $\mathbf{V} = \{w_1, \ldots, w_N\}$ of

*This work was funded by the SUMACC project supported by the French National Research Agency (ANR) under contract ANR-10-CORD-007.

Copyright is held by the author/owner(s).
MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

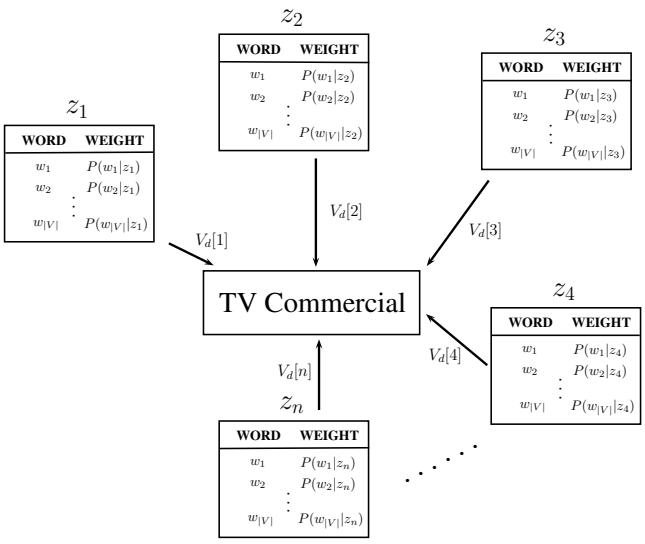

Figure 2: Mapping of a TV commercial in the topic space.

size N. This corpus contains $10,724$ Web pages related to brands of the commercials contained in D. This corpus is composed of $44,229,747$ words for a vocabulary of $4,476,153$ unique words. The topic representation is performed using a Latent Dirichlet Allocation (LDA) [1] approach. At the final LDA analysis, a topic space m of n topics is obtained with, for each theme z, the probability of each word w of v knowing z and for the entire model m, the probability of each theme z knowing the model m. Each TV commercial from both development and test set is mapped into the topic space (see figure 2).

2.2 Similarity measure

Each commercial have been mapped into the topic space to produce its vector representation. Then, commercials from the test set T that deal with the same subjects of commercials from the development set D are clustered. The cosine is used as a similarity measure:

$$cosine(V^d, V^t) = \alpha_{d,t} = \frac{\sum_{i=1}^{n} V^d[i] \times V^t[i]}{\sqrt{\sum_{i=1}^{n} V^d[i]^2}\sqrt{\sum_{i=1}^{n} V^t[i]^2}} \quad (2)$$

2.3 Rhythm pattern

The cosine measure, presented in the previous section, is also used to evaluate the similarity between a mean rhythm pattern vector $\overline{S^t}$ of a song and all the candidate songs S_k^t of the test set.

In details, each commercial from D, is related with a soundtrack that is represented with a rhythm pattern vector. In our experiments, the 10 rhythm features of the song are used (*speed*, *percussion*, *periodicity*, *rhythm pattern*...). As a result, each commercial is represented by a rhythm pattern vector of size 58. From the subset of soundtracks of the l nearest commercials from D, a mean rhythm vector $\overline{S}$ is performed as:

$$\overline{S} = \frac{1}{l}\sum_{d \in l} S^d \ .$$

Finally, the cosine measure between this mean rhythm $\overline{S}$ of the l nearest commercials from D and each commercial $(cosine(\overline{S}, S^t)_{t \in T})$ is used to find, from the soundtrack S^t of the test set T, the 5 songs from all the candidates having the closest rhythm pattern.

3. EXPERIMENTS AND RESULTS

The proposed system is evaluated in the MediaEval 2013 MusiClef benchmark [4]. The aim of this task is to predict for each video in the test set, the most suitable soundtrack from 5,000 candidate songs. The dataset is split into 3 sets. The development set contains multimodal information on 392 commercials (various metadata, Youtube uploader comments, various audio features, video features, web pages and text features). The test set is a set of 55 videos where a song should be associated using the recommandation set of 5,000 soundtracks (30 seconds long excerpts).

For each video in the test set, a ranked list of 5 candidate songs is proposed. The song prediction evaluation is manually performed using the Amazon Mechanical Turk platform. Three scores have been computed from our system output [4]:

- First rank average score: **2.16**
- Top 5 average score (arithmetic mean): **2.24**
- Top 5 average score (harmonic mean, taking rank into account): **2.22**

Considering that human judges rate the predicted songs from 1 (*very poor*) to 4 (*very well*), we can consider that our system is slightly better than the mean evaluation score (2) no matter the metric considered.

4. CONCLUSION

In this paper, an automatic system to assign a soundtrack to a TV commercial has been proposed. This system combines two media: textual commercial content and audio rhythm pattern.

5. REFERENCES

[1] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

[2] C. Bullerjahn. The effectiveness of music in television commercials. *Food Preferences and Taste: Continuity and Change*, 2:207, 1997.

[3] N. Hoeberichts. Music and advertising: The effect of music in television commercials on consumer attitudes. *Bachelor Thesis*, 2012.

[4] C. C. S. Liem, N. Orio, G. Peeters, and M. Scheld. MusiClef 2013: Soundtrack Selection for Commercials. In *MediaEval*, 2013.

[5] C. W. Park and S. M. Young. Consumer response to television commercials: The impact of involvement and background music on brand attitude formation. *Journal of Marketing Research*, pages 11–24, 1986.