

TuneSensor: A Semantic-Driven Music Recommendation Service For Digital Photo Albums

Jiansong Chao, Haofen Wang, Wenlei Zhou, Weinan Zhang, and Yong Yu

Department of Computer Science and Engineering
APEX Data & Knowledge Management Lab
Shanghai Jiao Tong University
Shanghai, P.R.China

{jiansong.chao,whfcarter,wenlei.zhouwl,wnzhang,yyu}@apex.sjtu.edu.cn

Abstract. Digital photo album softwares like iPhoto¹ have enjoyed great popularity for years. These years, online photo album services (e.g., Flickr² and Picasa³) have been becoming more and more popular with the development of social Web. In this paper, we demonstrate our effort called TuneSensor to recommend music for photo albums automatically. In particular, we exploit semantic data to represent both images and music. Furthermore, we leverage mining techniques to capture semantic relatedness between these different types of multimedia data, which is the essential step for recommendation.

1 Introduction

With the development of Web 2.0, digital photo album services have been becoming an indispensable part of social network sites. Representative examples such as Flickr and iPhoto are more and more popular among people. Besides photo publishing and sharing, they even allow users to assign music for some specified album. When browsing the photo album, it might be a fantastic experience if the background music matches the photos. However, the manual assignment limits the wide usage of such an attractive feature: (1) it will make a user exhausted if he has lots of albums; (2) it is hard for a user to select the suitable music if he has little related knowledge. Therefore, automatic background music recommendation for photo albums can greatly relief the user burden and improve the user experience. In this paper, we implemented a prototype demonstration called TuneSensor which is the first effort trying to recommend suitable music for the given photo album automatically. From the technical perspective, the main challenge of automatic recommendation lies in calculating the relatedness between music and images indicating whether they share the common artistic conception or express the similar emotion. In order to solve the above challenge,

¹ <http://www.apple.com/ilife/iphoto/>

² <http://www.flickr.com/>

³ <https://picasaweb.google.com/>



Fig. 1. Snapshot of TuneSensor User Interface.

we have to find a way to represent the very different data (i.e., image and music) in a unified semantic manner. With the advance of social Web, more and more multimedia data is annotated with tags. On the other hand, relatedness computing in the textual space has been well studied for years. Therefore, we make use of Flickr as the high-quality source to prepare for a large image database annotated with tags. AllMusic⁴ is used to associate mood tags to music. Further, we leverage WordNet⁵ [3] ontology to enrich these tags and disambiguate them into synsets so that we can easily connect images with their suitable music according to their emotional semantic relatedness. In order to recommend music for the input images, we exploit visual similarity between images so that the input images can be represented by several most similar images in our image database. The technical details can be found in the following sections. The snapshot of our demonstration is shown in Figure 1. Photos of the album are shown at the center and bottom. Recommended music is shown at the top right of the screen. Users can browse the album in slide view while listening to the music. In particular, this is a photo album of wedding dress. The playing background music recommended by TuneSensor is *Sugar Sugar* from *The Archies*.

2 Music Recommendation For Photo Albums

For an input photo album, TuneSensor analyses the album images, searches for the most related music in its music database, and finally recommends it for the input album as output. Here we introduce the system of TuneSensor via offline and online modules as shown in Figure 2.

⁴ <http://allmusic.com/>

⁵ <http://wordnet.princeton.edu/>

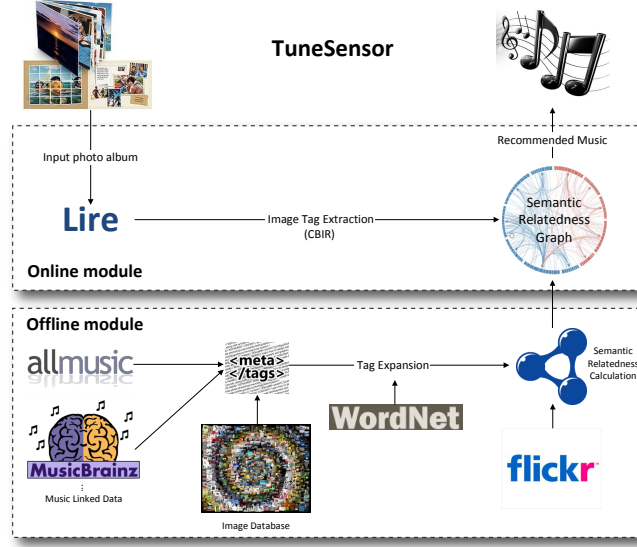


Fig. 2. The architecture of TuneSensor.

2.1 Construct Cross-media Semantic Relatedness Graph

The offline module constructs a cross-media semantic relatedness graph, which contains both image and music mood tag synsets as vertices, and the semantic relatedness between these synsets as edges. For music tag synset construction, we crawl metadata of each track from AllMusic. We also use service provided by MusicBrainz [2] to complement the information for each track. Moreover, we use WordNet to expand each music mood tag to a music tag vector so that the errors caused by ambiguous words and users' preference of words are avoided effectively. In this way, each track is represented by several mood tag vectors. For image tag synset construction, we choose WordNet synsets with typical subjective feelings and assign a certain number of images for each of these synsets. This construction is similar to a subset of ImageNet⁶. Similar with music synset expansion, each image tag synset can be converted into a semantic tag vector. With the vector representation of image and music synsets, we can get the relatedness between them by computing the relatedness between these two vectors by taking advantage of text semantic relatedness. In particular, we use the statistics of Flickr image tags. In this way, our cross-media semantic relatedness graph is constructed.

2.2 Music Recommendation Using Relatedness Graph

For a photo album as input, we use LIRE (Lucene Image Retrieval) [1] to compute the visual similarity between images in the album and the indexed synset

⁶ <http://www.image-net.org/>

images. In this way, we get the relatedness between input photo album and each semantic tag vector. Using the constructed semantic relatedness graph, each track is assigned with a relatedness score with the input album. Finally, the track with the highest match score is recommended.

3 Case Studies and Future Work

Table 1. Results of our music recommendation for photo albums.

Photo Album	Mood Tags Matched	Music Recommended	
	romantic, peaceful, dramatic, etc		“Because You Loved Me” by Celine Dion
	sensual, fun, sexy, etc		“Great Ball Of Fire” by Jerry Lee Lewis

Table 1 shows some case studies using TuneSensor for music recommendation. In this table, images in the first column are the photo albums as input. The related music mood tags found by TuneSensor are shown in the second column. The last column contains music recommended by TuneSensor and some related information. More recommendation cases can be experienced in our online demo <http://tunesensor.apexlab.org/>. In the experiment, our method achieved a performance of about 68% satisfaction measured by participants feedback. For the future work, we plan to scale up the number of synsets and the number of images for each synset. In this way, we can describe more visual content and feelings exactly. We can also provide TuneSensor as applications for photo album services like Flickr and Facebook album in the future. In addition, music search engine queried by images can be developed based on TuneSensor.

References

1. Lux Mathias, Savvas A. Chatzichristofis. Lire: Lucene Image Retrieval – An Extensible Java CBIR Library. In proceedings of the 16th ACM International Conference on Multimedia, pp. 1085-1088, Vancouver, Canada, 2008
2. Swartz A. MusicBrainz: A Semantic Web Service, in IEEE Intelligent Systems, Intelligent Web Services, January/February, 2002.
3. Christiane Fellbaum. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998