AN AUDIO AND MUSIC SIMILARITY AND RETRIEVAL SYSTEM BASED ON SPARSE FEATURE REPRESENTATIONS

Juhan Nam **Jorge Herrera**
CCRMA
Stanford University
{juhan,jorgeh}@ccrma.stanford.edu

Byeong-jun Han
School of Elec. Engr.
Korea University
hbj1147@korea.ac.kr

Kyogu Lee
MARG
Seoul National University
kglee@snu.ac.kr

ABSTRACT

In this paper, we present an audio and music similarity and retrieval(AMSR) system, which employed sparse feature representations. Our system first extracts Mel-scaled spectrum from audio snippet. Next, sparse restricted Boltzmann machine (RBM) is trained for feature representation. In order to enhance computational efficiency in retrieval phase, we employed locality sensitive hashing (LSH) method, which facilitates approximate nearest neighbor (ANN) search.

1. INTRODUCTION

Previous AMSR approaches have employed well-known acoustic and music features such as statistical analysis of magnitude spectrum, mel-frequency cepstral coefficients (MFCCs), chromagram and novelty curves. These features have been proven effective for most similarity computation and retrieval tasks, however, have to be re-analyzed for continuously changing reference audio and music datasets. Furthermore, finding new feature extraction methods has become time consuming and research laborious as well as needing more extremely professional knowledge about specific task.

This paper summarizes the submissions "AMSR_2012_1" and "AMSR_2012_2", which consist of following novel features: i) we employed feature learning concept so that our system extracts feature by learned feature extraction method. For this, sparse restricted Boltzmann machine (RBM) is included, and; ii) we applied locality sensitive hashing (LSH) in order to compare the performance with greedy search algorithm.

2. SPARSE FEATURE REPRESENTATIONS

Unlike previous hand-crafted feature extraction approaches such as MFCC, the proposed system first learns feature and then extracts feature by model. In order to learn and represent feature, we employed restricted Boltmann machine (RBM)[2]. All the features are sparse coded for efficient content representation. Meanwhile, we summarized song segments by performing max-pooling, which takes maximum value over pooled area.

3. APPROXIMATE NEAREST NEIGHBOR SEARCH

Like our previous approach[3] we compared the performances between approximate nearest neighbor (ANN) search and greedy search. Locality sensitive hashing (LSH) [4], a widely adopted ANN method in many retrieval research area, was employed.

4. CONCLUSIONS

This extended abstract introduces the details about our AMSR submissions. The most significant features of our submissions are: i) We employed sparse RBM for feature learning and representations, and; ii) We utilize LSH for ANN.

5. REFERENCES

[1] Juhan Nam, Jorge Herrera, Malcolm Slaney, Julius Smith, "Learning sparse feature representations for music annotation and retrieval," *Proceedings of the International Symposium on Music Information Retrieval (ISMIR 2012)*, 2012. (to be appeared)

[2] P. Smolensky, *Information processing in dynamical systems:Foundation of harmony theory*, MIT Press, Cambridge, 1986.

[3] Byeong-jun Han, Hyunwoo Kim, Ziwon Hyung, Kyogu Lee, Sheayun Lee, "A content-based music similarity retrieval scheme by using BoW representation and LSH-based retrieval," MIREX 2011: Audio and Music Similarity and Retrieval task.

[4] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," Proceedings of 30th Symposium on Theory of Computing, 1998.