CompositeMap: A Novel Music Similarity Measure for Personalized Multimodal Music Search

Bingjun Zhang[1], Qiaoliang Xiang[1], Ye Wang[1], Jialie Shen[2]
[1]School of Computing, National University of Singapore
[2]School of Information Systems, Singapore Management University
[1]{bingjun, xiangqiaoliang, wangye}@comp.nus.edu.sg
[2]jlshen@smu.edu.sg

ABSTRACT

How to measure and model the similarity between different music items is one of the most fundamental yet challenging research problems in music information retrieval. This paper demonstrates a novel multimodal and adaptive music similarity measure (CompositeMap) with its application in a personalized multimodal music search system. CompositeMap can effectively combine music properties from different aspects into compact signatures via supervised learning, which lays the foundation for effective and efficient music search. In addition, an incremental Locality Sensitive Hashing algorithm is developed to support more efficient search processes. Experimental results based on two large music collections reveal various advantages in effectiveness, efficiency, adaptiveness, and scalability of the proposed music similarity measure and the music search system.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation, Search process; H.5.5 [**Sound and Music Computing**]: Systems

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Music, Similarity Measure, Personalization, Search

1. INTRODUCTION

A music item consists of multiple aspects, such as title, genre, and mood. Each aspect contains rich semantics and its existing digital representations (low-level features, high-level concepts, etc.) are high-dimensional in nature. All these factors render measuring music similarity difficult.

The previous research on music similarity measure can be categorized as follows: 1) *metadata-based similarity measure* (MBSM) where text retrieval techniques are used to compare the similarity between the input keywords and the metadata around music items, e.g., music search in LastFm and YouTube; 2) *content-based similarity measure* (CBSM) where audio features are used as content descriptors to compute music similarity [2]; 3) *semantic description-based similarity measure* (SDSM) where each music item is annotated using a predefined vocabulary [4] and represented as a multinomial distribution which is used to measure the similarity.

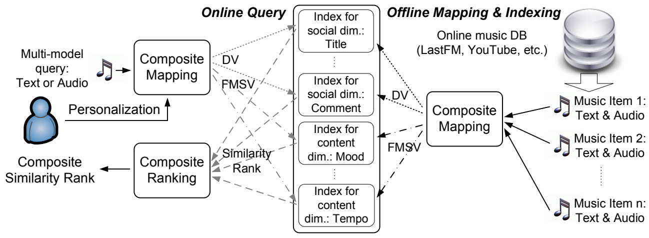

Online Query
Offline Mapping & Indexing
Multi-model query: Text or Audio
Composite Mapping
Personalization
Composite Similarity Rank
Composite Ranking
DV
FMSV
Similarity Rank
Index for social dim.: Title
Index for social dim.: Comment
Index for content dim.: Mood
Index for content dim.: Tempo
Online music DB (LastFM, YouTube, etc.)
DV
FMSV
Composite Mapping
Music Item 1: Text & Audio
Music Item 2: Text & Audio
Music Item n: Text & Audio

Figure 1: System architecture.

The disadvantages of the above music similarity measures are: it is expensive and difficult to represent multiple music aspects (timbre, melody, etc.) in human language (MBSM and SDSM); only a single music aspect or a holistic music meaning is considered in the similarity measure (CBSM), which restricts the adaptive and flexible nature of music semantics; the high computational complexity limits their practical uses in large databases (CBSM and SDSM).

To address the above problems, we propose a multifaceted music similarity measure called *CompositeMap*. Using CompositeMap, content-related dimensions (genre, mood, tempo, etc.) are modeled as Fuzzy Music Semantic Vectors (FMSVs) and social information-related dimensions (title, comments, etc.) are described as Document Vectors (DVs). Adaptive similarity between music items can be measured using any individual or combination of musical dimensions. This enables personalization to allow users to specify their preferred music dimensions in various search contexts. In addition, we also developed an indexing structure to further improve the efficiency of the retrieval process. Evaluation results demonstrate various advantages of the proposed music similarity measure and the multimodal music search system.

2. SYSTEM ARCHITECTURE

The system provides personalized multimodal music search. As shown in Fig. 1, it consists of the following components: CompositeMap which constructs music signatures of multiple music dimensions for both multimodal music queries and music items in the database; the indexing module for efficient and effective query of the music signature on each music dimension; CompositeRanking to combine the ranked lists from the personalized music dimensions into a more relevant one as the final search results.

MM'09, October 19–24, 2009, Beijing, China.
ACM 978-1-60558-608-3/09/10.

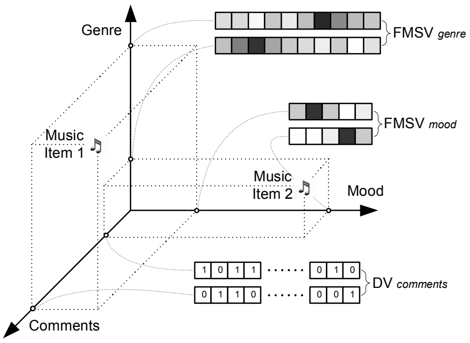


Figure 2: Conceptual illustration of CompositeMap.

3. SYSTEM IMPLEMENTATION

CompositeMap. We propose a compact music signature, called Fuzzy Music Semantic Vector (FMSV), to explicitly describe each music content-related dimension in a structured and human-understandable way [5]. By further representing the social information dimensions as Document Vectors (DVs) [3], a novel scheme called CompositeMap is proposed to map multiple and cross-modal music dimensions into a unified representation. A conceptual diagram of the CompositeMap is presented in Fig. 2. These music dimensions span a music space, in which adaptive music similarity can be measured between any two music items.

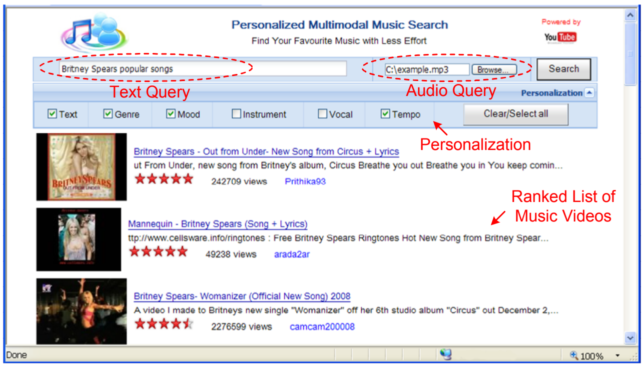
Incremental Locality Sensitive Hashing. Inspired by the inverted index in text retrieval, we develop a hybrid indexing framework to index each music dimension separately by its most suitable algorithm (Locality Sensitive Hashing [1] for FMSVs and inverted list for DVs) in order to build an overall efficient index for the whole music space.

CompositeRanking. Based on users' personalization of preferred music dimensions, nearest music items to the query are retrieved by iLSH in each of the personalized dimensions. The CompositeRanking then combines the multiple ranked lists with equal weights to form the final list as the search results.

4. EXPERIMENTS

By crawling the audio stream of music videos on YouTube, we built two test collections with 3020 and 100,000 music items respectively to evaluate the system effectiveness and efficiency. To simulate the realistic music search behavior, we designed music queries to simulate the personalization of any single music dimension or any combination of music dimensions. 24 subjects with different music backgrounds volunteered for the evaluation. We compared the CompositeMap with two baseline methods, i.e., audio features based (AF) and audio features plus principal component analysis based (AFPCA) similarity measures [5].

Experimental results show that CompositeMap clearly outperforms AF and AFPCA with statistically significant improvement measured by Precision@n. The average top-100 response time using CompositeMap is significantly less than using AF and AFPCA. As the data set gets larger, the response time of CompositeMap remains acceptable (≈ 0.5 seconds on the data set with 3000 samples and ≈ 1.7 seconds on the data set with 1 million simulated samples). Those results imply that music similarity measured by CompositeMap is more effective which leads to more relevant search results. With fast response time, the proposed framework scales well on large databases.



Britney Spears popular songs

C:\example.mp3

☑ Text ☑ Genre ☑ Mood ☐ Instrument ☐ Vocal ☑ Tempo

Britney Spears - Out from Under- New Song from Circus + Lyrics
ut From Under, new song from Britney's album, Circus Breathe you out Breathe you in You keep comin...
★★★★★ 242709 views Prithika93

Mannequin - Britney Spears (Song + Lyrics)
ttp://www.cellsware.info/ringtones : Free Britney Spears Ringtones Hot New Song from Britney Spear...
★★★★★ 49238 views arada2ar

Britney Spears- Womanizer (Official New Song) 2008
A video I made to Britneys new single "Womanizer" off her 6th studio album "Circus" out December 2,...
★★★★½ 2276599 views camcam200008

Figure 3: Search interface.

5. SYSTEM DEMONSTRATION

This demonstration illustrates the following advantages of the personalized multimodal music search system: superior effectiveness of music searches; better flexibility to accommodate various kinds of multimodal queries (both text and audio based); better adaptiveness to allow users to personalize preferred music dimensions in different search contexts.

The search interface is illustrated in Fig. 3. Users can input textual keywords and/or upload an audio clip to form a multimodal music query. The multimodal query increases the flexibility of query formation, which lessens the users' burden in forming meaningful queries. It also provides more information for the system to retrieve more relevant results. In addition, users can personalize their preferred music dimensions to search for their most desired music in different user contexts. This illustrates the adaptiveness of our system to better satisfy users' information need in various context with the same/different user queries.

6. CONCLUSION

In summary, we have presented CompositeMap, a novel multimodal music similarity measure, with its application in a personalized multimodal music search engine. We have evaluated the effectiveness, efficiency, adaptiveness and scalability of the system using two separate large scale music collections extracted from YouTube. The experimental results have shown the clear advantages of the proposed music similarity measure and its application in the search system.

7. ACKNOWLEDGEMENT

This work is supported by the research grant Multimodal Mobile Music Retrieval with WBS no. R-252-000-381-112.

8. REFERENCES

[1] A. Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *FOCS'06*, 2006.

[2] S. J. Downie. The music information retrieval evaluation exchange (2005 - 007): A window into music information retrieval research. *Acoustical Science and Technology*, 2008.

[3] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[4] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Towards musical query-by-semantic-description using the cal500 data set. In *Proc. of ACM SIGIR*, 2007.

[5] B. Zhang, J. Shen, Q. Xiang, and Y. Wang. Compositemap: A novel framework for music similarity measure. In *Proc. of ACM SIGIR*, 2009.