Chapter 3

Using Fuzzy Song Sets in Music Warehouses

François Deliège
Aalborg University, Denmark

Torben Bach Pedersen
Aalborg University, Denmark

ABSTRACT

The emergence of music recommendation systems calls for the development of new data management technologies able to query vast music collections. In this chapter, the authors present a music warehouse prototype able to perform efficient nearest neighbor searches in an arbitrary song similarity space. Using fuzzy song sets, the music warehouse offers a practical solution to three concrete musical data management scenarios: user musical preferences, user feedback, and song similarities. The authors investigate three practical approaches to tackle the storage issues of fuzzy song sets: tables, arrays, and compressed bitmaps. They confront theoretical estimates with practical implementation results and prove that, from a storage point of view, arrays and compressed bitmaps are both effective data structure solutions. With respect to speed, the authors show that operations on compressed bitmap offer a significant grain in performances for fuzzy song sets comprising a large number of songs. Finally, the authors argue that the presented results are not limited to music recommendations system but can be applied to other domains.

INTRODUCTION

Automatic music recommendation systems have recently gained tremendous popularity. To provide pertinent recommendations, music recommendation systems use fuzzy set theory (Zadeh, 1965) to combine user profiles, music features, and user feedback information. However, at the current growing speed, the database element of any recommendation system will soon become a bottleneck. Hence, appropriate musical data management tools, able to manipulate fuzzy sets and scale to large music collection...
and growing user communities, are needed. Music Warehouses (MWs) are dedicated data warehouses optimized for the storage and analysis of music content.

The contributions of this chapter are fourfold. First, based on a previous case study (Deliège & Pedersen, 2006), we propose three generic usage scenarios illustrating the current demands in musical data management. To answer these demands, we define fuzzy song sets and develop a query algebra for them. Second, to demonstrate the usefulness of fuzzy song sets, a prototypical MW composed of two multidimensional cubes is presented. Fuzzy song sets prove to be an adequate data representation to manipulate musical information. Third, we discuss three solutions for storing fuzzy song sets and fuzzy sets in general. We construct theoretical estimates for each storage solution. A practical implementation shows that the storage overhead represents a major part of the storage consumption and that two solutions are viable for large music collections. Fourth, we benchmark and compare the performance of the main operators previously presented for various sizes of both data structures. Experiments are conducted on a real music collection.

This chapter demonstrates how fuzzy set theory can be used in the context of music recommendation systems. All results presented in this chapter can be directly applied to standard fuzzy sets; the presented storage solutions remain generic and can thus be applied to a vast range of domains besides music recommendation and user preferences.

The remainder of this chapter is organized as follows. After presenting the related work on fuzzy sets for the management of musical data, we present three information scenarios that are commonly treated by music recommendation systems. We proceed by defining fuzzy song sets and an algebra. Two prototypical multidimensional cubes are presented; they illustrate the use of the algebra through query examples. Storage solutions are then discussed and precise storage estimates are proposed and experimentally validated. Next, a comparison of the performance of the fuzzy song set operators on the bitmap and array representations is conducted. Finally, we conclude and describe promising future research directions.

RELATED WORK

Research on music recommendation systems has received a lot of attention lately. Current trends on playlist generation are focused on how to improve recommendations based on user-specific constrains. For example, a playlist generator that learns music preferences by taking user feedback into account was presented by Pauws & Eggen (2001). Other new interesting approaches concentrate on aggregating different music features; for instance, Bosteels & Keere (2007) study the use of generalized conjunctions and disjunctions of fuzzy sets theory for combining audio similarity measures. However, fewer researchers have addressed the scalability issues raised by these methods in terms of storage and performance (Aucouturier & Pachet, 2002; Pampalk, 2005). This chapter focuses specifically on the storage and performance issues and proposes to manipulate a large collection of musical data where song similarities, user preferences and user feedbacks are represented with fuzzy sets.

A traditional database approach is to use a relational model such as the one proposed by Rubenstein that extends the entity-relationship data model to implement the notion of hierarchical ordering, commonly found in musical data (Rubenstein, 1987). A multimedia data model, following the layered model paradigm that consists of a data definition layer, a data manipulation layer, a data presentation layer, and a control layer, is presented by Wynblatt & Schloss (1995), but no query language is proposed. None
Related Content

A Dual PSO-Adaptive Mean Shift for Preprocessing Optimization on Degraded Document Images
[www.igi-global.com/article/a-dual-pso-adaptive-mean-shift-for-preprocessing-optimization-on-degraded-document-images/171463?camid=4v1a](www.igi-global.com/article/a-dual-pso-adaptive-mean-shift-for-preprocessing-optimization-on-degraded-document-images/171463?camid=4v1a)

Machine Learning for Texture Segmentation and Classification of Comic Image in SVG Compression

Vector Evaluated and Objective Switching Approaches of Artificial Bee Colony Algorithm (ABC) for Multi-Objective Design Optimization of Composite Plate Structures
[www.igi-global.com/chapter/vector-evaluated-objective-switching-approaches/69718?camid=4v1a](www.igi-global.com/chapter/vector-evaluated-objective-switching-approaches/69718?camid=4v1a)

Selection of Molybdenum-Filled Hip Implant Material Using Grey Relational Analysis Method