

Learning to Cope with Diversity in Music Retrieval

Thomas Mandl
Information Science, University of Hildesheim
Marienburger Platz 22
D-31141 Hildesheim – Germany
Tel.: ++49 - 5121 - 883 - 837
mandl@uni-hildesheim.de

Christa Womser-Hacker
Information Science, University of Hildesheim
Marienburger Platz 22
D-31141 Hildesheim – Germany
Tel.: ++49 - 5121 - 883 - 833
womser@uni-hildesheim.de

1. DIVERSITY IN MUSIC RETRIEVAL

Designers of music retrieval system are also confronted with a large number of choices. A look at some of the implementations proves this point. The forms for the user criteria include a significant number of different criteria.

The search for appropriate algorithms and representation atoms within the universe of possibilities seems to be a curse since each solution neglects several aspects. However, it can turn out to be a blessing when the diversity is integrated into a self adapting fusion system which considers many heterogeneous solutions. This paper presents a model including a machine learning approach to balance the influence of several system parameters according to users preferences.

The complexity of music as a formal system as well as a cultural phenomenon leads to difficulties for the computational representation. A single note by itself has no meaning and there is no correspondent to the syntax governing natural language which can be exploited in text retrieval. In addition, there exist many different technical formats for storing musical data.

Music as a cultural phenomenon has led to an abundance of musical styles. They all use different methods to express their intentions or let the listener enjoy a pleasing experience. Sounds, notes, pauses, tempi, instruments, and voices are combined in many ways. In this multimodal structure of music lies one of the roots of the problem of music retrieval. Depending on the style, different atoms are assembled for a composition. Which elements can be considered as prominent feature like a melody or a theme depends largely upon the style and the user need. Therefore, a formal model for the representation allows an optimization only for one style of music. With the possibility to store music in digital representations it has become increasingly attractive to search large music collections. As a matter of fact, the richness of music cannot be translated into all its dimensions (melody, harmony, rhythm etc.).

In contrast to texts, musical “documents” lack of separators necessary to identify semantic units like “words” or “phrases”. Like textual words, the same melodic pattern may occur in more than one piece of music, perhaps composed by different composers. Algorithms for stemming are necessary to detect variants and conflate them to the same stem. [1] have pointed out the issues of content-based indexing of musical data. The same entity can be represented in two different main forms: the notated and the acoustic form. Music communication is performed at two levels: the composer creates a musical structure while the interpreter (musician or singer) translates the written score into sounds. The resulting performances may differ a lot from each other. The information within a musical work can be identified at different levels: Melody, harmony, rhythm, and structure are

dimensions, carried out by written score, whereas in the case of musical performance other dimensions like timbre, articulation, and timing could be of interest.

The most widely spread mode for music retrieval is to search via similarity, whereas similarity in music retrieval presents several difficulties: what part of a song is likely to be perceived as the theme of the music?

Most systems seek for similarity on a level of pitch. Usually these systems like SEMEX [2] only process monophonic melodies, however, for some musical styles polyphonic matching would be desirable. A global representation might e.g. only consider a histogram analysis of pitch values. Approaches from speech recognition have also been applied to music retrieval [3].

The parameters for music retrieval discussed above need to be considered when implementing a system. Each parameter represents one dimension in the solution space for a specific retrieval system. The space of potential solutions is highly dimensional. The search for a solution within the high dimensional space has the goal of achieving a good retrieval quality. Therefore, the search is guided either by heuristics or by empirical results. Finding an optimal solution requires a large testbed of tasks and evaluation of the results by users or experts.

However, when the conditions change a different solution might be optimal. These changes may be the consequence of different queries, new user interests or changes of the music content. Now, another solution might produce the optimal result.

2. THE MIMOR MODEL

Fusion of various approaches is widely used in computer science. The goal of applying several algorithms is to improve the overall performance. Fusion methods delegate a task to several systems and integrate their results into one final result presented to the user. Ideally, the weaknesses of one method do not have a large negative influence on the final result because they are superimposed by another method. A typical example are committee machines in machine learning [4]. The fusion may be implemented as a voting scheme or as a weighted linear

MIMOR (Multiple Indexing and Method-Object Relations) is a fusion approach taking advantage of heterogeneity [5, 6]. The MIMOR model samples users’ relevance feedback to predict optimal method-object relations where methods are indexing algorithms or retrieval models. These are assigned to the characteristics of users and documents with the goal of improving the overall retrieval quality. From a computational viewpoint, MIMOR is designed as a linear combination of the results of different retrieval systems. The contribution of each system or algorithm to the fusion result is governed by a weight for that system.

$$RSV_{MIMOR}(doc_i) = \frac{\sum_{system=1}^N (\omega_{system} RSV_{system}(doc_i))}{N}$$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2002 IRCAM – Centre Pompidou

A central aspect in MIMOR is learning. The weight of the linear combination of each information retrieval system is adapted according to the success of the system measured by the relevance feedback of the users. A system which gave a high retrieval status value (RSV) and consequently a high rank to a document which then received positive relevance feedback should be able to contribute with a higher weight to the final result. The following formula enables such a learning process:

$$\omega_{system} = \varepsilon RF_{user}(doc_i) RSV_{system}(doc_i)$$

ε learning rate

Learning in MIMOR leads to a fusion which combines the individual systems in an optimal way. As a result, MIMOR takes advantage of two of the most promising strategies for improving information retrieval systems. These are relevance feedback and fusion. However, the optimal combination may depend on the context and especially on the users' individual perspectives and the characteristics of the documents. Therefore, MIMOR needs to consider context.

The performance of information retrieval systems differs from domain to domain. Characteristics of the documents relevant for the indexing procedure may be responsible for that. MIMOR builds upon the idea that formal properties can be exploited to improve fusion. Some retrieval methods work better e.g. for short documents. The weight of these systems should be high for short documents only.

The properties are modeled as clusters. All documents which have a property in common belong to the same cluster. Each cluster can develop its own adequate MIMOR model with weights for all participating systems.

The term clustering is usually used for non-supervised learning methods which find structures in data without hypotheses. However, the assignment of music pieces to clusters for the improvement of information retrieval processes may also be carried out with supervised learning methods. Therefore, the term cluster in this abstract does not restrict this process to algorithms based on unsupervised learning. Both supervised learning methods for pre-defined classes and even human assignment are compatible with MIMOR.

3. M-MIMOR: SELF ADAPTATION FOR MUSIC RETRIEVAL SYSTEMS

The MIMOR approach is very well suited for music retrieval. Music retrieval incorporates high diversity along several dimensions of system parameters. The choice of parameter values is almost arbitrary. On the other hand, MIMOR offers a fusion method which learns from the preferences of the user. The M-MIMOR approach makes productive use of the multidimensionality of music retrieval. It integrates heterogeneous poly-representation into a self adapting system. The different perspectives of users can be expressed by relevance feedback and serve as direction for a learning process which ultimately leads to an optimal solution for a user within a certain context. Instead of focusing on one value for each system parameter, each user receives the most adequate mixture of the options available.

Genre detection systems have been developed for music [7]. Therefore, genre can be used as one feature in M-MIMOR. The calculation of the similarity between query and musical objects

needs to consider not only the systems involved. In addition, the clusters to which an object belongs and the membership function M enter the formal model.

$$RSV_{MIMOR}(doc_i) = \frac{\sum_{sys=1}^N \sum_{clus=1}^K (\omega_{sys,clus} \eta_{clus} RSV_{sys}(doc_i))}{K \cdot N}$$

$$M(doc_i) = (\eta_{cluster1}, \eta_{cluster2}, \eta_{cluster3}, \dots)$$

4. CONCLUSION

This article introduces a model for music retrieval which automatically learns to adapt itself to the cognitive preferences of the user and supports the multimodal nature of music. Since the evaluation of musical objects is highly subjective, a retrieval system needs to dynamically identify the most adequate combination of system parameters for the user. M-MIMOR manages this integration in a linear combination of many possible variables. Consequently, M-MIMOR takes the personalization and adaptivity one step further.

As a result, no viewpoint expressed in a certain algorithm or representation method needs to be neglected but may contribute with the proper weight to the final result. The fusion of a diversity of perspectives will ultimately lead to better retrieval performance.

MIMOR has been implemented in JAVA and has been tested in text retrieval. Improvements against individual systems could be measured. The integration on music retrieval tools based on JAVA could be easily realized.

5. REFERENCES

- [1] Melucci, M. and Orio, N. Musical Information Retrieval using Melodic Surface. Proc Conf. Digital Libraries 1999, Berkeley, CA.
- [2] Lemström, K. and Perttu, S. SEMEX – An Efficient Music Retrieval Prototype. Intl Symp for Music Information Retrieval. ISMIR 2000. <URL:http://ciir.cs.umass.edu/music2000>
- [3] Logan, B. Mel Frequency Cepstral Coefficients for Music Modeling. Intl Symp for Music Information Retrieval. ISMIR 2000. <URL:http://ciir.cs.umass.edu/music2000>
- [4] Haykin, S. Neural Networks: A Comprehensive Foundation. 1999.
- [5] Womser-Hacker, C. Das MIMOR-Modell. Mehrfachindexierung zur dynamischen Methoden-Objekt-Relationierung im Information Retrieval. Habilitationsschrift. Universität Regensburg, Informationswissenschaft. 1997.
- [6] Mandl, T. and Womser-Hacker, C. Probability Based Clustering for Document and User Properties. In: Ojala, T. (ed.). Infotech Oulo International Workshop on Information Retrieval (IR 2001). Oulu, Finland. 2001. pp. 100-107.
- [7] Tzanetakis, G. Essl, G. and Cook, P. Automatic musical genre classification of audio signals. Intl Symp for Music Information Retrieval. ISMIR 2001. <URL:http://ismir2001.indiana.edu>