User-Adaptive Music Information Retrieval

Sebastian Stober, Andreas Nürnberger

Automatic structuring is one means to ease access to large music collections – be it for organisation or exploration. The AUCOMA project (Adaptive User-Centered Organization of Music Archives) aims to find ways to make such a structuring intuitively understandable to a user through automatic adaptation. This article describes the motivation of the project, discusses related work in the field of music information retrieval and presents first project results.

1 Introduction

One of the big challenges of computer science in the 21st century is the digital media explosion. Steadily growing hard-drives are filled with personal media collections comprising e.g. music, photos and videos. With increasing collection size maintenance becomes a more and more tedious task, but without manual organization effort it gets harder to access specific pieces of media or even to keep an overview. Typically, a large portion of the digital content is just "collecting dust" because the user has simply forgotten about it. Here, computer science and especially artificial intelligence can help to improve awareness and accessibility of such data: Automatic structuring is one means to ease access to media collections, be it for organization or for exploration. Moreover, users would greatly benefit if a system would not just structure the collection for easier access but would structure it in a way that is intuitively understandable for the individual user by adapting to personal preferences and needs. Unfortunately, such aspects of individualization have been only a minor issue of research in the field of Multimedia Retrieval. At best, interfaces for media collection access allow for adaption by the user. However, they are largely lacking the ability to learn from user actions and to adapt on their own without explicit intervention of the user. The aim of the AUCOMA project is to develop intuitive, non-obstrusive, user-adaptive methods for media collection access with special focus on music information retrieval (MIR).

Dealing with music data, the following considerations serve as motivation for the project: Firstly, music can be described by a large variety of facets comprising e.g. simple tags (artist, title etc.), content-based features ranging from simple loudness to complex timbre descriptions, harmonics, meters and tempi, instrumentation, and lyrics but also information about the production and publishing process as well as the general reception in the public expressed in reviews or chart positions. This diversity of features makes music especially interesting from the data mining point of view and allows to transfer results to different domains. Secondly, perception of music is highly subjective and may depend on a person's background. A musician, for instance, might especially look after structures, harmonics or instrumentation (possibly paying – conscious- or unconsciously – special attention to his own instrument). Non-musicians will perhaps focus more on overall timbre or general mood. Others, in turn, may have a high interest in the lyrics as long as they are able to understand the particular language. Finally and most im-

portantly, music can be considered an integral part of daily life even though it may often only play a background role. There may be common contexts in which music is consumed as well as contexts that are particular to an individual listener. Either way, the choice of music listened to in each context is supposed to be highly individual. The large variety of usage contexts makes MIR especially interesting for research in the area of user modelling and personalization. Given these considerations, the project's approach to user-adaptation is two-fold:

1. Starting from various features that describe music as mentioned above, a complex multi-facet similarity measure is constructed where a similarity facet is computed either on a single feature or a combination of features. Weighting each similarity facet allows for adaptability. A weighting scheme then represents a user's preference for grouping similar songs together and can be applied for any similarity-based structuring approach. The user is not asked explicitly to adjust the weighting to fit his needs – most likely this would be a very difficult thing to do anyway and some users might not even be aware of such preferences. Instead, the weighting has to be learned by watching the user interact with the collection. Section 2 gives details on this aspect.

2. Another way of adaptation is to utilize knowledge about personal listening habits. The idea is to infer idiosyncratic genres that represent different listening contexts. Such information can then either be used directly to browse the collection or to enrich a similarity-based structuring as orientation aid or as separate facet. Also, a combined context- and similarity-based structuring might be possible. Details on this aspect are discussed in Section 3.

2 User-Specific Similarity Measure

2.1 Adaptable and Adaptive MIR Systems

The idea of adapting similarity measures is not new: MPeer [4] allows to adjust the weight of three facets of music description in a similarity measure through an intuitive joystick interface for finding a set of similar songs given an anchor song. The facets comprise the audio content, the lyrics and cultural metadata collected from the web. From a study with 10 users, it was concluded that users tend to use nearly similar joystick settings throughout different environments. Though the joystick

interface is very intuitive, it is unclear whether it may be applied to more than 3 facets. Similarly, the E-Mu Jukebox [21] allows changing the similarity function that is applied to create a playlist from a seed song. Here, five similarity components (sound, tempo, mood, genre and year) are visually represented by adapters that can be dragged on a bull's eye. The closer a component is to the center, the higher is its weight in the similarity computation. This interface is scalable with respect to the number of facets but less intuitive. It may be hard for a user to explicitly specify a weighting scheme for the facets as this is usually something that only subconsciously exists. Especially with an increasing number of facets this is likely to become more difficult. Indeed, a user study with 22 participants showed that the users found the system harder to use but at the same time more useful compared to two control systems.

In contrast to the former systems, PATS (Personalized Automatic Track Selection) [16] is an adapative system for playlist generation that does not require manual adjustment of the underlying similarity measure but learns from user feedback. The system generates a playlist for a specific user context through dynamic clustering. The user can then select songs in the playlist that in his opinion do not fit to the current context-of-use. From this preference feedback, new feature weights in the underlying similarity measure are derived by an inductive learning algorithm based on the construction of a decision tree that uncovers the feature values classifying songs into the categories "preferred" and "rejected". Similarly, the system described in [22] uses machine learning techniques for user-adaptive playlists generation. Though the basic idea to learn from user feedback is indeed very similar to the approach taken in the AUCOMA project, the usage scenario and the adaption algorithm are completely different: Systems for automatic playlist generation aim to compile lists of similar songs given one or more seed songs. In contrast to that, the goal for AUCOMA is to structure a whole collection of songs. Hence, such a binary classification does not suffice.

Another adaptive system for playlist generation called PAPA (Physiology and Purpose-Aware Automatic Playlist Generation) [14] as well as the already commercially available BODiBEAT music player[1] uses sensors that measure several bio-signals (such as the pulse) of the user as immediate feedback for the music currently played. This information is then used to learn which characteristics of music have a certain effect on the user. Based on this, continuously adapting model playlists for different purposes can be created. Though this method of getting immediate feedback for continuous adaptation is highly interesting, it is not applicable in the usage scenario of our approach.

2.2 Preliminary Results

A first prototype of a user-adaptive system for structuring and exploring music collections was presented in [19]. It is based on a multi-facet similarity measure with about 20 facets covering e.g. sound, harmonics, lyrics and information about the production process. The system was tested on a corpus containing 282 songs of The Beatles. Though this corpus is rather small and homogeneous compared to other collections cointaining songs of several artists and genres, it had the advantage that high-quality feature data was already widely available such as the manually
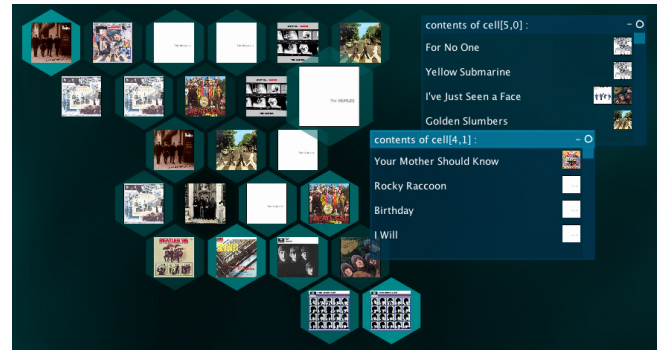

Figure 1: Screenshot of the prototype: grid and two cell content windows.

annotated chord labels provided by the Queen Mary University, London.

Using the initially unbiased multi-facet similarity measure, a growing self-organizing map is induced as described in [13] clustering similar song of the music collection into hexagonal cluster cells. The result is a two-dimensional topology that preserves the neighborhood relations of the originally high dimensional feature space, i.e. not only songs within the same cluster cell are similar to each other but also songs of cells in the neighborhood are expected to be more similar than those in more distant cells. Using a growing approach ensures that only as many cells are created as are actually needed. Further, the approach is incremental: Songs may be added to the collection without having to relearn the whole map from scratch – instead, it is only extended.

The cells of the generated hexagonal grid can be seen as "virtual folders", each one containing a set of similar songs. A screenshot of the prototype user interface is shown in Fig. 1.[2] Cells are labelled with the album cover(s) that are most frequently linked with the songs contained in the cell. Clicking on a cell opens a window that displays its content. For each song, the title and (if available) the album covers are listed. The user may change the location of songs on the map by simple drag-and-drop actions. Each movement of a song causes a change in the underlying multi-facet similarity measure based on a quadratic optimization scheme introduced in [20]. As a result, the location of other songs is modified as well. First experiments simulating user interaction with the system showed that during such a stepwise adaptation the similarity measure indeed converges to one that captures the user's preferences.

2.3 Current and Future Work

Having verified the general approach, the goal is now to extend the prototype to be able to process large-scale collections containing tens of thousands of songs. One important aspect is the automatic extraction of features replacing manual annotation. As a first step, an automatic chord recognizer based on the approach described in [17] and an extractor for meta-data from wikipedia will be integrated. Further, the flat clustering of the self-organizing map does not scale well with the collection size. Therefore, the clustering approach is currently extended to

---

[1]http://www.yamaha.com/bodibeat/

[2]A demo video of the prototype user interface is available at http://www.findke.ovgu.de/aucoma

build a hierarchical structure. Here, approaches will be studied that have been already successfully applied to text collections [1, 2]. Additionally, alternative visualization techniques will be investigated.

3 Learning Idiosyncratic Genres

3.1 Motivation

Objective genre classification appears to be a hard task and a general applicable classification scheme which holds for everybody has not yet been agreed upon. This problem has been bothering the MIR research community for about a decade. However, several studies indicate that there might be meaningful user-specific genres emerging from usage patterns that people consciously or unconsciously use when they access music collections or describe music:

In a user study [10] that analyzed organization and access techniques for personal music collections, several "idiosyncratic genres", i.e. "genres" peculiar to the individual, could be identified that users tend to use to classify and organize their music. These idiosyncratic genres comply largely with the usage context. Typical examples could be "music for driving (and keeping me awake)", "music for programming" or "music to relax in the evening after a long working day". Further, an analysis of requests at the answering service "Google Answers"[3] in the category "music" [3] revealed that such descriptions were also used in this public setting. In a larger survey [12] on search strategies for public music retrieval systems, more than 40% stated that they would query or browse by usage context if this would be supported by the system. Strong correlations between genre, artist, album and the usage context also became apparent in a more recent study [9].

Even for the construction of an objective taxonomy of 378 music genres [15], features referring to consumer context ("audience location") and usage ("danceability") were important criteria. Such meta-data is already exploited in a commercial application [8] to select music for a desired atmosphere in hotels, restaurants and cafes. However, the respective properties need to be assigned manually by experts and if necessary can only be adapted by hand.

It would be very desirable to have at least a semi-automatic context assignment from automatically retrievable or measurable data. According to the definition by Dey [6], any information that can be used to characterise the situation of a person, place or object of consideration makes up its context. He differentiates four types of primary context: location, identity, time, and activity [7]. In the music information retrieval domain, there exists already a variety of systems that capture time, (user) identity and location, e.g. the Audioscrobbler[4] plug-in from last.fm[5]. However, this information is rarely used to describe a usage context and to the authors' knowledge it has so far not been used for personalized access to music collections. Recalling the phenomenon of idiosyncratic genres, this yields a high potential for supporting an individual user in maintaining and using his personal music collection.

3.2 Preliminary Results

For logging the listening context information together with the played songs, a plug-in for the foobar2000, Winamp and iTunes music player was developed. Whenever a song is played, the plug-in records its ID3 metadata together with a time stamp, the location and local weather information obtained via a web-service. In a small pilot study with 8 participants, usage data was collected over a time period of about three month. The results were presented in [11] together with a basic prototype interface for browsing a music collection by usage context and a broad discussion on what and how further context information may be collected. Although, for significant results, data collected over a longer time period would be required, it could be demonstrated how widely available environmental data can be exploited to allow organization, structuring and exploration of music collections by personal listening contexts.

For a prototype context browser the elastic list technique [18] that was developed for browsing multi-facetted data structures was adopted. This approach enhances traditional facet browsing interfaces such as presented in [5] for music collections that allow a user to explore a data set by filtering available metadata information. Additionally to facet browser filtering, elastic lists can also display value distributions proportionally and visualize unusualness.[6]

3.3 Current and Future Work

Amongst the various possibilities for extended context logging that emerged during the pilot study, logging the current activity at the computer (i.e. the programs used) and measuring of the background noise level between two songs through a (commonly built-in) microphone appear to be the most promising. Both approaches require no special hardware and promise useful context information. However, clearly privacy is the most important issue here. So the question is rather not what is technically possible but how much information about his activities a user is willing to share. As more sophisticated methods come closer to surveillance, the user must be fully informed about the extend of the collected data and in full control of whether he wants this data to be logged or not. Most importantly, it needs to be proven that this additional information is indeed helpful, i.e. the user has a benefit from providing this information.

Currently, a study is prepared that aims to investigate the general acceptance for context logging and to which extend average users would be willing to provide data. Based on the results of this study, the logger will be extended respectively. Further insights into the privacy issue might also arise from analysis of the context data collected during the "urbansync" self-eperiment of Stephan Baumann.[7]

4 Summary and Conclusions

The AUCOMA research project follows a two-fold approach towards user-adaptive music information retrieval: One focus lies on finding ways to automatically learn a similarity measure that

---

[3]http://answers.google.com/answers/
[4]http://www.audioscrobbler.net/
[5]http://last.fm/

[6]An online demo of the original elastic lists user interface can be found at http://well-formed-data.net/experiments/elastic_lists/
[7]The logged data and information on the urbansync project is available at http://urbansync.wordpress.com/

captures the way the user compares songs and can be applied for any similarity-based structuring approach. The other is to derive idiosyncratic genres from usage context information that can automatically be recorded. This paper motivated both approaches, discussed related work and presented some early project results. Apart from further developing the two approaches towards user-adaptation separately, the upcoming challenge lies in bringing both together.

5 Acknowledgements

References

[1] K. Bade, M. Hermkes, and A. Nürnberger. User oriented hierarchical information organization and retrieval. In *Proc. of 2007 European Conference on Machine Learning (ECML'07)*, 2007.

[2] K. Bade and A. Nürnberger. Creating a cluster hierarchy under constraints of a partially known hierarchy. In *Proc. of 2008 SIAM International Conference on Data Mining (SDM'08)*, 2008.

[3] D. Bainbridge, S. J. Cunningham, and J. S. Downie. How people describe their music information needs: A grounded theory analysis of music queries. In *Proc. of 4th Intl. Conf. on Music Information Retrieval (ISMIR'03)*, 2003.

[4] S. Baumann and J. Halloran. An ecological approach to multimodal subjective music similarity perception. In *Proc. of 1st Conf. on Interdisciplinary Musicology (CIM'04)*, 2004.

[5] R. Dachselt and M. Frisch. Mambo: a facet-based zoomable music browser. In *Proc. of 6th Intl. Conf. on Mobile and Ubiquitous Multimedia (MUM'07)*, 2007.

[6] A. K. Dey. Understanding and using context. *Personal Ubiquitous Computing*, 5(1):4–7, 2001.

[7] A. K. Dey and G. D. Abowd. Towards a better understanding of context and context-awareness. In *Computer Human Interaction (CHI) 2000 Workshop on the What, Who, Where, When, Why and How of Context-Awareness*, 2000.

[8] S. Govaerts, N. Corthaut, and E. Duval. Moody tunes: the rockanango project. In *7th Intl. Conf. on Music Information Retrieval (ISMIR'06)*, 2006.

[9] X. Hu, J. S. Downie, and A. Ehmann. Exploiting recommended usage metadata: Exploratory analyses. In *7th Intl. Conf. on Music Information Retrieval (ISMIR'06)*, 2006.

[10] S. Jones, S. J. Cunningham, and M. Jones. Organizing digital music for use: an examination of personal music collections. In *Proc. of 5th Intl. Conf. on Music Information Retrieval (ISMIR'04)*, 2004.

[11] V. Laube, C. Moewes, and S. Stober. Browsing music by usage context. In *Proc. of 2nd Workshop on Learning the Semantics of Audio Signals (LSAS)*, 2008.

[12] J. H. Lee and J. S. Downie. Survey of music information needs, uses, and seeking behaviours: Preliminary findings. In *Proc. of 5th Intl. Conf. on Music Information Retrieval (ISMIR'04)*, 2004.

[13] A. Nürnberger and A. Klose. Improving clustering and visualization of multimedia data using interactive user feedback. In *Proc. of 9th Intl. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'02)*, 2002.

[14] N. Oliver and L. Kreger-Stickles. PAPA: Physiology and purpose-aware automatic playlist generation. In *7th Intl. Conf. on Music Information Retrieval (ISMIR'06)*, 2006.

[15] F. Pachet and D. Cazaly. A taxonomy of musical genres. In *Proc. of Content-Based Multimedia Information Access (RIAO)*, 2000.

[16] S. Pauws and B. Eggen. PATS: Realization and user evaluation of an automatic playlist generator. In *Proc. of 3rd Intl. Conf. on Music Information Retrieval (ISMIR'02)*, 2002.

[17] J. Reinhard, S. Stober, and A. Nürnberger. Enhancing chord classification through neighbourhood histograms. In *Proc. of 6th Intl. Workshop on Content-Based Multimedia Indexing (CBMI'08)*, 2008.

[18] M. Stefaner and B. Muller. Elastic lists for facet browsers. In *Proc. of 18th Intl. Conf. on Database and Expert Systems Applications (DEXA'07)*, 2007.

[19] S. Stober and A. Nürnberger. Towards user-adaptive structuring and organization of music collections. In *Proc. of 6th Intl. Workshop on Adaptive Multimedia Retrieval (AMR'08)*, 2008.

[20] S. Stober and A. Nürnberger. User modelling for interactive user-adaptive collection structuring. In *Proc. of 5th Intl. Workshop on Adaptive Multimedia Retrieval (AMR'07)*. Springer, 2008.

[21] F. Vignoli and S. Pauws. A music retrieval system based on user driven similarity and its evaluation. In *Proc. of 6th Intl. Conf. on Music Information Retrieval (ISMIR'05)*, 2005.

[22] K. Wolter, C. Bastuck, and D. Gärtner. Adaptive user modeling for content-based music retrieval. In *Proc. of 6th Intl. Workshop on Adaptive Multimedia Retrieval (AMR'08)*, 2008.

Contact

Data & Knowledge Engineering Group
Faculty of Computer Science, Otto-von-Guericke-University
Universitätsplatz 1, 39106 Magdeburg
Tel.: +49 (0)391 67-18287
Fax: +49 (0)391 67-12020
Email: sebastian.stober@ovgu.de
Email: andreas.nuernberger@ovgu.de
Homepage: http://www.findke.ovgu.de/

Bild | **Sebastian Stober** is Ph.D-student at the Data & Knowledge Engineering Group of the OvGU Magdeburg. He received his diploma degree in computer science in 2005. His main research interests are machine learning, music information retrieval and adaptive systems.

Bild | **Andreas Nürnberger** studied computer science and economics at TU Braunschweig. After he received his Ph.D. in computer science from the OvGU Magdeburg in 2001 he moved as PostDoc to the University of California at Berkeley where he worked on adaptive soft computing and visualisation techniques for information retrieval systems. Since 2003 he was Juniorprofessor for Information Retrieval at OvGU and headed the DFG funded Emmy Noether junior research group. Since 2007 he is professor for 'Data and Knowledge Engineering' at the OvGU.